

University of Nottingham

UK | CHINA | MALAYSIA

Human Induced Pluripotent Stem Cell Models of Glycosaminoglycan Associated Diseases

by

Emma Louise Barker, BSc (Hons)

A doctoral thesis submitted to The University of Nottingham for the degree of Doctor of Philosophy

March 2023

Acknowledgements

I would first like to thank my supervisor, Professor Cathy Merry, for all the guidance, knowledge and support she has provided to me over the years. Thank you for allowing me to understand the beauty of glycosaminoglycans and pushing me to be the scientist I have become today. I would also like to thank Dr Sara Pijuan-Galitó for always being around, steering me in the right direction after an extremely difficult lockdown and teaching me everything I know about cell culture. Special appreciation to Dr Mattéa Finelli, who joined my supervisory teams latterly; thank you for challenging me on all things molecular biology related and helping me with all things neural. It is fair to say that I could not have achieved any of this without the invaluable help and support of these three people.

I would also like to acknowledge Dr Alexander Kondrashov for being a gene editing wizard and knowing everything there is to know. Thank you for sharing your knowledge with me and for always being there when it wasn't going quite as I expected.

I cannot go without mentioning all of my friends in the BDI, they have no idea how much they have helped me through the latter stages of my PhD by always being around for a chat and never turning down a pub Friday. Special thanks to the all the past and present members of Stem Cell Glyco for supporting me through my time with you. Again, I could not have done this without them.

Finally, I would like to acknowledge the two most important people in my life, my mum and dad. I cannot begin to describe how grateful I am to both of you for the unwavering love and support you have provided me with, even though I have made it difficult at times! You have both been with me through every step of this PhD and I appreciate it more than anything. Thank you for being my greatest inspirations in life.

'The whole of science is nothing more than a refinement of everyday thinking' – Albert Einstein

Table of Contents

List of Figures 1
List of Tables 3
Abbreviations 4
Abstract 7
Declaration of Collaborative Work 8
Publications 9
COVID-19 Impact Statement 10

Chapter 1 12
1.1 Thesis Overview 12
1.2 Proteoglycans and Glycosaminoglycans 13
1.2.1 The Extracellular Matrix 13
1.2.2 Proteoglycans 14
1.2.3 Glycosaminoglycans 17
1.3 Heparan Sulphate Biosynthesis 20
1.3.1 Tetrasaccharide Formation 20
1.3.2 Initiation and Elongation 20
1.3.3 N-Deacetylation and N-sulphation of Glucosamine 25
1.3.4 Epimerization of Glucuronic Acids 28
1.3.5 2-O-Sulphation 29
1.3.5.1 Heparan Sulphate 2-O-Sulphation and Kidney Development 31
1.3.6 6-O-Sulphation 33
1.3.7 3-O-Sulphation 35
1.3.8 Post Synthetic Modifications 37
1.4 Heparan Sulphate Biosynthesis and Disease 38
1.5 Heparan Sulphate Degradation 42
1.5.1 The Mucopolysaccharidoses 45
1.5.2 Mucopolysaccharidosis Type IIIA 46
1.5.2.1 Treatment of Mucopolysaccharidosis Type IIIA 48
1.5.2.2 Human Induced Pluripotent Stem Cell Models of Mucopolysaccharidosis Type IIIA 51
1.6 Methods of Glycosaminoglycan Analysis 54
1.6.1 Compositional Disaccharide Analysis 54
1.6.2 Metabolic Radiolabelling .. **Error! Bookmark not defined.**
1.6.3 Glycosaminoglycans Antibodies 57
1.6.4 Time of Flight Secondary Ion Mass Spectrometry 58
1.7 Human Induced Pluripotent Stem Cells 59

1.7.1 Human Induced Pluripotent Stem Cells for Disease Modelling 60
1.7.2 Heparan Sulphate and Human Induced Pluripotent Stem Cells 62
1.7.3 Challenges with Human Induced Pluripotent Stem Cells 66
1.8 CRISPR/Cas9 Gene Editing 66
1.8.1 Methods of CRISPR/Cas9 Delivery 68
1.8.2 CRISPR/Cas9 and Glycosaminoglycans 71
1.9 Project Summary 71

Chapter 2 73
2.1 Human Induced Pluripotent Stem Cell Culture 73
2.1.1 Vitronectin Coating 73
2.1.2 Human Induced Pluripotent Stem Cell Maintenance 73
2.2 Immunostaining 74
2.3 Gene Expression Analysis 76

Chapter 3 77
3.1 Introduction 77
3.1.1 Chapter Overview 77
3.1.2 CRISPR/Cas9 Targeting of SGSH 77
3.2 Materials and Methods 79
3.2.1 gRNA Design 79
3.2.2 Generation of Arms of Homology 79
3.2.3 Targeting Plasmid Generation 80
3.3 Results 82
3.3.1 Developing Homology Directed Repair Plasmids 82
3.3.1.1 Determining Optimal PCR Conditions 82
3.3.1.2 Digesting Arms of Homology 82
3.3.1.3 Ligating Arms of Homology 83
3.4 Discussion 85
3.4.1 Is Plasmid Delivery the Correct Choice For CRISPR/Cas9 Knockout of SGSH? 85
3.4.2 Future Work 86

Chapter 4 88
4.1 Introduction 88
4.1.1 Chapter Overview 88
4.1.2 CRISPR/Cas9 Targeting of HS2ST1 88
4.2 Materials and Methods 90
4.2.1 Fibroblast Cell Culture 90

4.2.2 Extraction of Glycosaminoglycans from Conditioned Medium and Cell Lysates 90
4.2.3 Purification of AMAC Fluorescent Label 90
4.2.4 Preparation of AMAC-labelled Disaccharides 91
4.2.5 Preparation and Analysis of Disaccharide Standards For RP-HPLC 91
4.2.6 Chinese Hamster Ovary Cell Culture 92
4.2.7 crRNA Design 92
4.2.8 gRNA Duplex and RNP Complex Formation 93
4.2.9 Transfection and Picking 93
4.2.10 Genotyping 94
4.2.11 Flow Cytometry 95
4.3 Results 97
4.3.1 The Heparan Sulphate Disaccharide Composition from Patients with Defects in *HS2ST1* Is Significantly Different When Compared to Control Cell Lines 97
4.3.2 HS2ST1 Targeting in REBLPAT hiPSCs 100
4.3.3 HS2ST1 mRNA Is Still Present in All Clones 102
4.3.4 Heparan Sulphate 2-O-Sulphation is Undetectable in Knockout Cell Lines 103
4.3.5 hiPSC Lines with Reduced 2-O-Sulphation Have Increased 6-O-Sulphation 108
4.4 Discussion 111
4.4.1 A Comparison of Mouse Models and Human Patients .. 111
4.4.2 Implications of Loss of HS2ST1 on Humans 113
4.4.3 Are hiPSCs That Don't Display 2-O-Sulphation Knockouts?. 115
4.4.4 Summary of CRISPR/Cas9 Targeting 117

Chapter 5 119
5.1 Introduction 120
5.1.1 Chapter Overview 120
5.1.2 Definitive Endoderm Differentiation 121
5.1.3 Ectoderm (Neural) Differentiation 122
5.2 Materials and Methods 125
5.2.1 Definitive Endoderm Differentiation 125
5.2.2 Definitive Endoderm Flow Cytometry 125
5.2.3 Neural Precursor Cell Differentiation 126
5.2.4 Neural Precursor Cell to Neuron Differentiation 126
5.3 Results 128

5.3.1 hiPSCs With Functionally Altered HS2ST1 Express Pluripotency Markers 128
5.3.2 hiPSCs with Functionally Altered HS2ST1 Can Differentiate to Definitive Endoderm 130
5.3.3 hiPSCs with Functionally Altered HS2ST1 Can Differentiate to Neural Precursor Cells 135
5.3.4 Optimizing A Protocol for Neural Differentiation 144
5.3.5 HS2ST1$^{-/-}$ hiPSCs Display Reduced FGFR3 Levels 146
5.4 Discussion 147
5.4.1 Altered Heparan Sulphate Structure Does Not Impair Definitive Endoderm and Neural Precursor Cell Formation 147
5.4.2 HS2ST1$^{-/-}$ hiPSCs Display Decreased Expression of FGFR3 149
5.4.3 How Are hiPSCs With Functionally Altered HS2ST1 Maintaining Pluripotency and Differentiating? 150
5.4.4 Summary of Characterization of hiPSCs With Altered HS2ST1 151

Chapter 6 152
6.1 Final Discussion and Conclusions 152
6.2 Future Work 154
References 159
Appendix 178

List of Figures

Figure 1 – Proteoglycans 16
Figure 2 – Glycosaminoglycan Disaccharide Composition 19
Figure 3 - Tetrasaccharide Linker Synthesis 21
Figure 4 - Heparan Sulphate Biosynthesis 24
Figure 5 - HS2ST1 Patient Mutations 41
Figure 6 - Heparan Sulphate Degradation 44
Figure 7 – Mutations of MPS IIIA Patient Derived hiPSCs 51
Figure 8 – Compositional GAG Disaccharide Analysis by AMAC-labelling RP-HPLC 56
Figure 9 - Hybridoma Derived GAG Antibodies 57
Figure 10 - Pluripotency Signalling in hiPSCs 64
Figure 11 - Differentiation Signalling in hiPSCs 65
Figure 12 - Mechanisms of DNA Repair After Cas9 Double Stranded Break 68
Figure 13 - Methods of CRISPR/Cas9 Machinery Delivery 70
Figure 14 - Targeting Strategy to Disrupt SGSH 78
Figure 15 – Agarose Gels of Arms of Homology 84
Figure 16 – Heparan Sulphate Disaccharide Composition of Patient 1 98
Figure 17 – Heparan Sulphate Disaccharide Composition of Serum Containing Medium 99
Figure 18 - CRISPR Targeting of HS2ST1 101
Figure 19 - HS2ST1 mRNA Expression 102
Figure 20 – Hs2st1$^{-/-}$ CHO Cell Lines Produce Heparan Sulphate 103
Figure 21 – Genetically Altered hiPSCs Produce Heparan Sulphate 104
Figure 22 – Knockout Confirmation Using Flow Cytometry 105
Figure 23 – Hs2st1$^{-/-}$ CHO Cells Do Not Stain for AO4B08 106
Figure 24 – Knockout Confirmation Using AO4B08 and Immunocytochemistry 107
Figure 25 – Heparan Sulphate Disaccharide Composition of Cell Lysates 109
Figure 26 – Heparan Sulphate Disaccharide Composition of Conditioned Medium 110
Figure 27 - HS2ST1 CRISPR/Cas9 Summary 119
Figure 28 - Summary of Differentiation Protocols 124
Figure 29 - HS2ST1$^{+/-}$ and HS2ST1$^{-/-}$ hiPSCs Express Pluripotency Markers 129

Figure 30 - Definitive Endoderm Differentiation 130
Figure 31 - hiPSCs with Altered HS2ST1 Function Differentiate to Definitive Endoderm.....131
Figure 32 – Definitive Endoderm Differentiation (NANOG, OCT4; REBLPAT WT, HS2ST1$^{+/-}$, HS2ST1$^{-/-}$).....132
Figure 33 – Definitive Endoderm Differentiation (FOXA2; REBLPAT WT, HS2ST1$^{+/-}$, HS2ST1$^{-/-}$).....133
Figure 34 - Definitive Endoderm Differentiation (SOX17; REBLPAT WT, HS2ST1$^{+/-}$, HS2ST1$^{-/-}$) 134
Figure 35 - Optimizing Seeding Densities for Neural Precursor Cell Differentiation.....135
Figure 36 - Neural Progenitor Cells Form in REBLPAT WT hiPSC After 7 Days. 136
Figure 37 - Neural Precursor Cell Differentiation 137
Figure 38 - Neural Precursor Cell Differentiation (NANOG, OCT4; REBLPAT WT, HS2ST1$^{+/-}$).....139
Figure 39 - Neural Precursor Cell Differentiation (NANOG, OCT4; HS2ST1$^{-/-}$).....140
Figure 40 - Neural Precursor Cell Differentiation (NESTIN, SOX2; REBLPAT WT, HS2ST1$^{+/-}$) 141
Figure 41 - Neural Precursor Cell Differentiation (NESTIN, SOX2; HS2ST1$^{-/-}$).....142
Figure 42 - Neural Precursor Cell Differentiation (PAX6; REBLPAT WT, HS2ST1$^{+/-}$) 143
Figure 43 – Neural Precursor Cell Differentiation (PAX6; HS2ST1$^{-/-}$) 144
Figure 44 - Neural Differentiation 145
Figure 45 - HS2ST1$^{-/-}$ hiPSCs Show Reduced FGFR3 Expression146
Figure 46 - Summary of Work and Future Work.....159

List of Tables

Table 1 - HS2ST1 Patient Mutations..... 40
Table 2 - The Mucopolysaccharidoses..... 46
Table 3 - Heparan Sulphate Phage Display Derived ScFv Antibodies 58
Table 4 - Cell Culture Volumes..... 74
Table 5 - Immunostaining Antibody Details..... 75
Table 6 - Primer Sequences Used for qRT-PCR 76
Table 7 - Primer Sequences for gRNA and Arms of Homology 79
Table 8 - PCR Cycling Conditions and Annealing Gradient..... 80
Table 9 - Restriction Enzyme Details. 81
Table 10 - crRNa Sequences..... 92
Table 11 - Genotyping Primer Sequences 95
Table 12 - PCR Cycling Conditions..... 95
Table 13 - Antibodies Used For Knockout Confirmation via Flow Cytometry 96
Table 14 - Antibodies Used for Definitive Endoderm Flow Cytometry. 126

Abbreviations

AAV	adeno-associated virus
AMAC	2-aminocridone
BBB	blood brain barrier
BMP	bone morphogenetic protein
BSA	bovine serum albumin
Cas9	CRISPR associated protein 9
CHO	Chinese hamster ovary
CNS	central nervous system
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
CS	chondroitin sulphate
DE	definitive endoderm
DMEM	Dulbecco's Modified Eagle Medium
DPBS	Dulbecco's phosphate buffered saline
DS	dermatan sulphate
DSB	double stranded break
DVL	dishevelled
E8	Essential 8
ECM	extracellular matrix
EDS	Ehlers-Danlos syndrome
EHS	Engelbreth Holm Swarm
ER	endoplasmic reticulum
ERK	extracellular signal-receptor kinase
ERT	enzyme replacement therapy
EXT	exostosin glycosyltransferase
EXTL	exostosin like glycosyltransferase
FACS	fluorescence-activated cell sorting
FGF	fibroblast growth factor
FGFR	fibroblast growth factor receptor
Fz	frizzled
GAG	glycosaminoglycan
Gal	galactose
GalNAc	N-acetyl galactosamine
GDNF	glial cell derived neurotrophic factor
GDS	glucuronate 2-sulphatase
GlcA	glucuronic acid
GLCE	HS C5 epimerase
GlcNAc	N-acetyl glucosamine
GlcNS	N-sulphated glucosamine

GPCR	G-protein coupled receptor
GPI	glycosylphosphatidylinositol
gRNA	guide RNA
GSK3β	glycogen synthase kinase 3β
GUSB	β-glucuronidase
HA	hyaluronan
HAS	hyaluronan synthase
HDR	homology directed repair
hESC	human embryonic stem cell
HGSNAT	heparan-α-glucosaminide N-acetyltransferase
Hh	hedgehog
HiPSC	human induced pluripotent stem cell
HS	heparan sulphate
HS2ST	heparan sulphate 2-O-sulphotransferase
HS3ST	heparan sulphate 3-O-sulphotransferase
HS6ST	heparan sulphate 6-O-sulphotransferase
HSC	haematopoietic stem cells
HSPG	heparan sulphate proteoglycans
IdoA	iduronic acid
IDS	iduronate 2-sulphatase
IDUA	α-L-iduronase
KO	knockout
KS	keratan sulphate
LSD	lysosomal storage disorder
MAPK	mitogen activated protein kinase
MEF	mouse embryonic fibroblasts
mESCs	mouse embryonic stem cells
MLD	metachromatic leukodystrophy
MPS	mucopolysaccharidosis
MS	mass spectronetry
MSC	mesenchymal stem cells
NAGLU	α-N-acetylglucosaminidase
NDST	N-deacetylase/N-sulphotransferase
NHEJ	non homologous end joining
NPC	neural precursor cells
NSC	neural stem cell
PAM	protospacer adjacent motif
PAPS	3’-phosphoadenosine 5’-phosphosulfate
PG	proteoglycan
PI3K/AKT	phosphatidylinositol 3 kinase/AKT
PSC	pluripotent stem cell

qRT-PCR	quantitative real time polymerase chain reaction
RNP	ribonucleoprotein
RP-HPLC	reverse phase high performance liquid chromatography
ScFv	single-chain variable fragments
SGSH	N-sulphoglucosamine sulphohydrolase
SLRP	small leucine rich proteoglycan
SRT	substrate reduction therapy
TALENS	transcription activator-like effector nucleases
TGFβ	transforming growth factor-β
ToF-SIMS	time of flight secondary ion mass spectrometry
tracrRNA	transacting CRISPR RNA
UA	uronic acid
VTN-N	vitronectin
WT	wild type
Xyl	xylose
XYLT	xylosyltransferase
ZFN	zinc finger nucleases

Abstract

Heparan sulphate (HS) is a key component of the extracellular matrix and regulates critical biological processes. It is a long chain, linear polysaccharides composed of repeating uronic acid (UA) and N-acetyl glucosamine (GlcNAc) disaccharide units. Sulphated regions within HS chains control their location, function and affinity for specific ligands. The sulphated domains occur via the stepwise action of several enzymes, N-deacetylase/N-sulfotransferases (NDSTs), C5-epimerase (GLCE), HS 2-O-sulphotransferase (HS2ST1), HS 6-O-sulphotransferases (HS6STs), and HS 3-O-sulphotransferases (HS3STs). HS2ST1 acts on the UA residues of HS while the NDSTs, HS6STs and HS3STs act on the glucosamine. The enzymes work in a step wise manner, and it is generally accepted they work in the sequential manner described above. Previously, lack of the HS2ST1 enzyme in mice was shown to be embryonically lethal due to skeletal abnormalities and kidney agenesis. *Hs2st1*$^{-/-}$ mouse HS lacks 2-O-sulphation, has increased 6-O-sulphation and N-sulphation, and therefore retains an approximately normal sulphation distribution allowing for fibroblast growth factor 2 (FGF2) to still elicit a response, despite reduced binding affinity. Recently, four human patients with diverse genetic alterations in the *HS2ST1* gene have been reported and have skeletal and renal abnormalities, including absent kidneys. In this doctoral thesis, HS was isolated from patient fibroblasts and the disaccharide composition was analysed revealing a similar composition to *Hs2st1*$^{-/-}$ mice including a loss of 2-O-sulphation and an increase in N-sulphation and 6-O-sulphation. *HS2ST1* was knocked out in human induced pluripotent stem cells (hiPSCs) using CRISPR/Cas9. Confirmation of knockout, and loss of 2-O-sulphation was carried out by flow cytometry and immunostaining using a HS antibody (10E4) and an antibody that preferentially binds the 2-O-sulphation modification (AO4B08). HS2ST1$^{-/-}$ and HS2ST1$^{+/-}$ hiPSCs were differentiated to neural precursor cells and definitive endoderm and immunostaining, flow cytometry and quantitative reverse transcription PCR (qRT-PCR) were used to assess the differentiation potential of the cells. Initial studies show that knockout cell lines are able to maintain their pluripotency and differentiate but further work is required to assess alterations in their signalling pathways and terminal differentiation capacity.

Declaration of Collaborative Work

Elements of this doctoral thesis were conducted in collaboration with other research groups. All collaborations were carried out and discussed with the author. Key collaborations have been noted in the text and have also been listed below:

- Section 4.2.1 – Pauline Schneeberger and Professor Kerstin Kutsche (Institute of Human Genetics, Hamburg) identified patients with mutations in *HS2ST1*. Pauline Schneeberger carried out fibroblast cell culture and collected conditioned medium and cell lysates. The author took over analysis from this point and prepared samples and analysed the data.

- Section 4.2.7 to 4.2.10 – Dr. Alexander Kondrashov provided valuable knowledge on the ribonucleoprotein CRISPR/Cas9 targeting strategy and also aided with understanding the sequencing results. Guide design, targeting and knockout confirmation were all carried out by the author.

- Section 5.2.1 & 5.2.2 – Liam Reed (University of Nottingham, UK) provided support with the endoderm differentiation protocol and flow cytometry analysis. The actual differentiation and analysis were carried out by the author.

- Section 5.2.3 & 5.2.4 – Dr. Mattéa Finelli (University of Nottingham, UK) helped with the generation of a suitable neural differentiation protocol and guided the author by checking for neural morphologies. The differentiation and subsequent analyses were carried out by the author.

Publications

- Schneeberger, P.E., von Elsner, L., Barker, E.L., Meinecke, P., Marquardt, I., Alawi, M., Steindl, K., Joset, P., Rauch, A., Zwijnenburg, P.J., Weiss, M.M., Merry, C.L.R., and Kutsche, K., 2020. Bi-allelic pathogenic variants in HS2ST1 cause a syndrome characterized by developmental delay and corpus callosum, skeletal, and renal abnormalities. *The American Journal of Human Genetics,* 107(6), pp.1044-1061.
    - The author carried out the HS compositional disaccharide analysis and subsequent figure generation. The author helped with the preparation of all areas of the manuscript where HS is discussed.

COVID-19 Impact Statement

The COVID-19 pandemic started during the second year of my PhD, though I was still in the first year of working on my project. Our research group had just moved into a new building that required a lab shut down for four months in order to move all of our equipment to the new facility. There were several delays in getting set up due to unavoidable hold ups in the safety office. During the building move, we were advised to undertake any work that we could do away from labs (such as writing and data analysis) meaning that during the pandemic, most of what I could have done was already completed. We had only been up and running as a fully functioning laboratory for a short period before the pandemic started.

During the COVID-19 lockdown, there was no lab access for six months. Upon returning to the labs, there were severe supply chain issues with plasticware, pipette tips and essential reagents having a lead time of up to three months, and as a result we were advised to limit our work. In addition, the building had imposed a capacity limit which, due to high demand, meant the time I could spend in the lab was extremely limited. Furthermore, during the initial stages of returning to work, all face-to-face contact was prohibited meaning training requiring close working was not possible. Although I had learnt many of the techniques I needed during my first year, I had not had the chance to be trained on any of the equipment I would be using and so experienced further delays outside of my control while I was waiting for safety signoffs and risk assessments. Furthermore, issues with equipment such as failure, servicing and demand could not be mitigated and have all delayed my project progressing. Due to these delays, I was unable to generate my knockout cell lines until 2021 with the consequence that work planned for the later stages of my PhD could not be carried out. Although we were only unable to access labs for six months, our building was not up and running to full capacity until early 2022 meaning the delays and problems I experienced were compounded. In addition to this, I also had to undertake an industrial placement as part of my funding requirements, which took a further 3 months out of my lab access time.

Just prior to the pandemic, I obtained some data that I was able to analyse and publish (with collaborators) during lockdown. The outcome of this data resulted in my project focus being switched. This provided me with a lengthy time to plan my CRISPR/Cas9 targeting and subsequent experiments (although the supply and training issues forced further changes). To mitigate the effects of the lockdown a different method of CRISPR/Cas9 targeting was chosen that would allow me to begin targeting immediately upon return to work. Unlike with the initial project plan, the CRISPR/Cas9 delivery did not require any in-house synthesis meaning no molecular biology was required prior to targeting the cells, saving time ultimately. Changes in my supervisory team were made to ensure I had all the support I needed following lockdown. This helped to make sure I had the best outcome from my lab work, though the effects of the lockdown are still evident throughout my PhD.

Chapter 1

Introduction

1.1 Thesis Overview

This doctoral project was carried out as part of the BBSRC Doctoral Training programme at the University of Nottingham. The project began investigating mucopolysaccharidosis (MPS) type IIIA, a disease associated with heparan sulphate (HS) degradation but changed its focus to HS biosynthesis. The overall aim of the project was to understand how glycosaminoglycans (GAGs) and development are altered in a novel disease associated with the GAG biosynthetic pathway. The work shown here builds on previous work carried out in mice and uses human induced pluripotent stem cells (hiPSCs) to achieve the overall aim. This thesis is structured as follows:

Chapter 1 introduces the extracellular matrix (ECM), proteoglycans (PGs) and GAGs. The biosynthesis and degradation of HS are discussed in detail and the clinical implications of alterations in these pathways are also mentioned. Additionally, the roles of HS in development are introduced and the effects of HS abnormalities on developmental process are reviewed. Potential uses for hiPSCs, and the challenges associated with these are presented, while the role of HS in hiPSC signalling is also discussed. This chapter finishes with an introduction to clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated protein 9 (Cas9), more commonly denoted as CRISPR/Cas9.

Chapter 2 is a methods chapter that details methods commonly used throughout this thesis.

Chapters 3-5 are data chapters, each with a brief introduction, a materials and methods section containing methods specific to that chapter, a results section and a discussion.

Chapter 3 discusses the initial work involved in creating a CRISPR/Cas9 knockout (KO) of N-sulphoglucosamine sulphohydrolase (*SGSH*), an HS degradation associated gene. Mutations in *SGSH* result in a disease known as MPS IIIA. The work in this chapter was halted due to COVID-19 and new discoveries changed the focus of the project.

Chapter 4 presents data on the HS composition of a patient discovered to have a novel mutation in heparan sulphate 2-O-sulphotransferase (*HS2ST1*), an HS biosynthetic associated gene. Following this discovery, a CRISPR/Cas9 KO of *HS2ST1* was generated in hiPSCs and KO and subsequent loss of 2-O-sulphation was confirmed.

Chapter 5 describes the phenotypic characterization of the hiPSCs generated in Chapter 4. The pluripotency status of the hiPSCs is determined and they are differentiated to endoderm and ectoderm to identify whether *HS2ST1* mutation has any effect on developmental pathways.

Chapter 6 draws overall main conclusions from the data chapters and discusses them in relation to recent literature. Future work is also discussed in this chapter.

1.2 Proteoglycans and Glycosaminoglycans

1.2.1 The Extracellular Matrix

The extracellular matrix (ECM) is a dynamic environment that surrounds cells and provides an essential physical scaffold, but also provides biochemical and biomechanical support (Walma and Yamada, 2020). The ECM is also responsible for binding growth factors and interacting with cell receptors to elicit downstream signalling and increase or decrease gene transcription. The macromolecules that make up the ECM can be usefully segregated into three specific classes: fibrous proteins, glycoproteins and proteoglycans (PGs). The exact ECM composition is highly tissue specific and also varies based on age and disease-state. For example, the ECM that surrounds the basement membrane is very rich in perlecan, an HSPG that provides the ECM with

characteristics that aid with glomerular filtration (Frantz, Stewart and Weaver, 2010). Furthermore, the ECM surrounding the CNS is very rich in chondroitin sulphate proteoglycans (CSPGs) that help with regeneration and plasticity. Additionally, these CSPG can interact with other ECM components, including hyaluronan, to form aggregates known as perineuronal nets (Fawcett *et al.*, 2022). The ECM is highly regulated and disruptions in its composition can lead to, or exacerbate, a diseased state. The ECM surrounding cancer tissues is abnormally deposited and remodelled which increases the stiffness of the tissue.

Fibrous proteins consist primarily of collagen while glycoproteins include fibronectin and laminin. Collagen is responsible for much of the structural support, cell adhesion and play a role in chemotaxis. The other ECM proteins also have mechanical roles; elastin is responsible for the ability of tissues to stretch, and fibronectin is involved in ECM organization and cell attachment (Frantz, Stewart and Weaver, 2010; Hoshiba *et al.*, 2016). The PG portion of the ECM can have an effect on the mechanical properties by sequestering water or forming aggregates, thereby altering such characteristics (de Mattos Pimenta Vidal *et al.*, 2017; Silver, Kelkar and Deshmukh, 2021). Importantly, the PGs (or their GAG side chains) are responsible for binding growth factors and act as a reservoir within the ECM to generate morphogen or growth factor gradients (Hynes, 2009)

1.2.2 Proteoglycans

PGs are proteins with covalently bonded GAG chains. PGs are ubiquitously expressed both at the plasma membrane and in the ECM (**Figure 1**). PGs are found in all multicellular organisms including the simpler *Drosophila Melanogaster,* although these organisms contain less PG isoforms compared to more complex multicellular organisms. For example, they only contain two glypican isoforms and one syndecan isoform compared to the six and four (respectively) found in humans. Furthermore, PGs have also been shown to be present in ancient multicellular organisms such as *Hydra* although the core proteins are yet to be characterized (Esko and Lindahl, 2001). PGs can be separated into three different classes based on their location, cell surface PGs (e.g. glypicans and syndecans), extracellular PGs (e.g. aggrecan

and biglycan) and intracellular PGs (e.g. serglycin). Different core proteins carry specific GAG chains meaning side chains are not interchangeable. **Figure 1** shows glypican bound to HS side chains and this PG cannot bind other GAGs in place of HS. PGs have varying numbers of GAG attachment sites ranging from one on decorin to more than one hundred in aggrecan (Roughley and Mort, 2014; Velleman and Song, 2017).

When PGs are covalently bonded to HS, they are known as heparan sulphate proteoglycans (HSPGs). HSPGs can be further split into three different families: the glypican family (glypican 1-6) which are glycosyl-phosphatidylinositol (GPI) anchored to the cell membrane, the syndecan family which are transmembrane HSPGs and extracellular matrix HSPGs (e.g. perlecan and testican). The six members of the glypican family can be separated into two different groups – glypicans 1,2,4 and 6 in one family and glypicans 3 and 5 in another. The groups are based on similar amino-acid identity with around 25% similarity within each group (Filmus, Capurro and Rast, 2009a). Glypicans are around 60-70kDa in size and have 14 conserved cysteine residues which is characteristic of these HSPGs. The cysteine rich residues are located in the central domain with the HS attachment sites being serine residues between the same central domain and the GPI-anchor (Fransson, 2003). The HS chains are located very close to the cell surface which supports the suggestion that a key function of glypicans is to interact with other cell-surface molecules (Filmus, Capurro and Rast, 2009b). Although the precise function of glypicans is not well known, it is clear that they play a vital role during development with observed functions in the regulation of Wnt, fibroblast growth factor (FGF) and Hedgehog (Hh) signalling (De Cat and David, 2001). Glypicans are acted on by an extracellular lipase, Notum, which releases the PG from its GPI-anchor into the ECM; this process is known as shedding (Filmus, Capurro and Rast, 2009a). It is important to note that other membrane bound PGs are also shed in a similar way to change the cell surface and alter the ligand binding capacity of the cells and therefore the intrinsic signalling pathways that are activated or inhibited (Nam and Park, 2012).

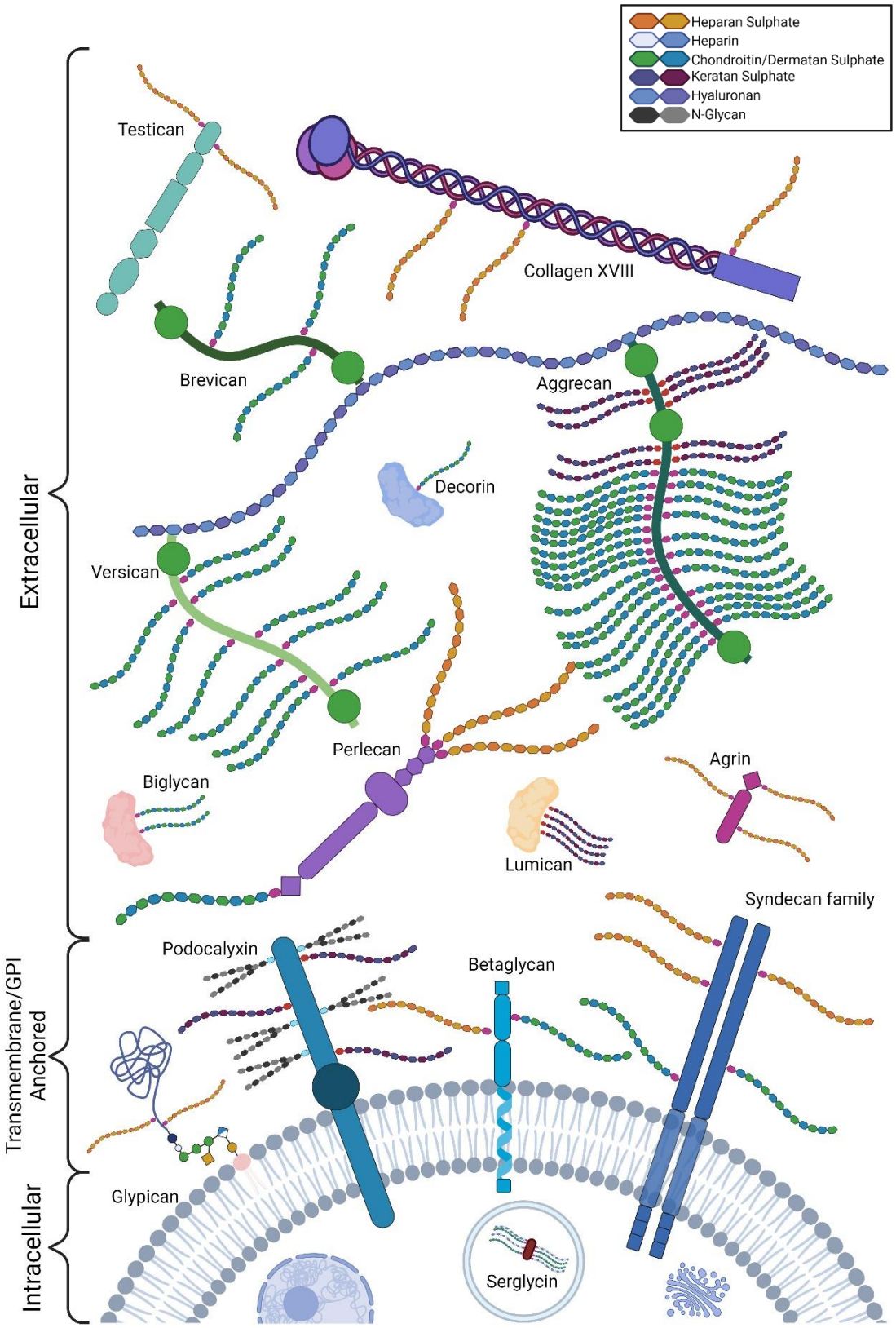


Figure 1 – Proteoglycans. Small leucine rich proteins (SLRPs; biglycan, lumican or decorin) and HA binding PGs such as aggrecan, versican and brevican are found in the ECM. Perlecan, agrin and collagen XVIII are all HSPGs which are found in the pericellular matrix. The syndecan and glypican families along with betaglycan and podocalyxin are membrane bound PGs. Serglycin is found inside mast cells.

Syndecans are structurally different from glypicans in that they are type I transmembrane proteins. They are made up of an ectodomain that contains the GAG attachments sites, a hydrophobic transmembrane domain and a C-terminal cytoplasmic domain which is usually short. The cytoplasmic domains consist of two highly conserved regions that flank a variable region that is responsible for syndecan-specific signalling (Carey, 1997; Gondelaud and Ricard-Blum, 2019). Although classified as transmembrane, syndecans can also act as soluble HSPGs via the shedding of their ectodomain. It is thought that the shedding regulates the amount of HSPGs both on the cell surface and in the ECM and therefore plays a role in cell function and cell signalling (Iozzo and Schaefer, 2015). There are four different syndecans present in vertebrates which are then further divided into two subfamilies – syndecan 1 and syndecan 3 form one subfamily while syndecan 2 and syndecan 4 form another. All four of the syndecans have HS attachment sites on the extracellular domains whereas only syndecan 1 and syndecan 3 have chondroitin sulphate attachment sites which sit closer to the plasma membrane (Tanabe, Yokota and Kohgo, 1999; Stepp *et al.*, 2015). Syndecans act as coreceptors for a multitude of HS binding growth factors. They also play an important role in cell-cell and cell-ECM interactions (Tkachenko, Rhodes and Simons, 2005).

Extracellular HSPGs include perlecan, agrin, collagen XVIII and collagen XV. Perlecan and agrin share homology at the C-terminus and both have roles in cardiovascular and musculoskeletal system regulation. Perlecan is the largest of the extracellular HSPGs and due to its widespread distribution, has a large variety of functions including roles in cell adhesion, cell death and angiogenesis. The SPOCK family (previously known as testican) are also extracellular HSPGs that can be found in the central nervous system (Iozzo and Schaefer, 2015).

1.2.3 Glycosaminoglycans

Glycosaminoglycans (GAGs) are long, linear heteropolysaccharides that carry a negative charge. They are composed of repeating disaccharide units and, with the exception of hyaluronan (HA), can

be sulphated at various positions (**Figure 2**). The various sulphation positions result in a myriad of disaccharides that make up the sulphated GAG chain. **Figure 2** shows how the sulphation patterning of the disaccharides results in a highly complex molecule.

Most GAGs are made up of repeating disaccharide units of a uronic acid (UA; glucuronic acid, GlcA; iduronic acid, IdoA) and an amino sugar (N-acetylglucosamine, GlcNAc; N-acetylgalactosamine, GalNAc). However, keratan sulphate (KS) lacks the uronic acid which is substituted for galactose (Gal). GAGs are classified into five families based on their disaccharide composition: heparan sulphate (HS) and heparin (GlcA/IdoA-GlcNAc), chondroitin sulphate (CS; GlcA-GalNAc), dermatan sulphate (DS; GlcA/IdoA-GalNAc), keratan sulphate (Gal-GlcNAc) and hyaluronan (GlcA-GlcNAc) (Souza-Fernandes, Pelosi and Rocco, 2006; Merry *et al.*, 2022). The sulphated GAGs are covalently attached to core proteins, known as PGs (**Figure 1**). The sulphation pattern of the GAGs is extremely important because this determines its ligand binding ability and therefore its function (Souza-Fernandes, Pelosi and Rocco, 2006).

All sulphated GAGs are synthesized in the Golgi apparatus but are attached to the core protein in different ways via the different linker regions seen in **Figure 2** (Merry *et al.*, 2022). HA is synthesized by three hyaluronan synthases (HASs) which act inside the cell at the plasma membrane (Hascall, Weigel and Toole, 2016).

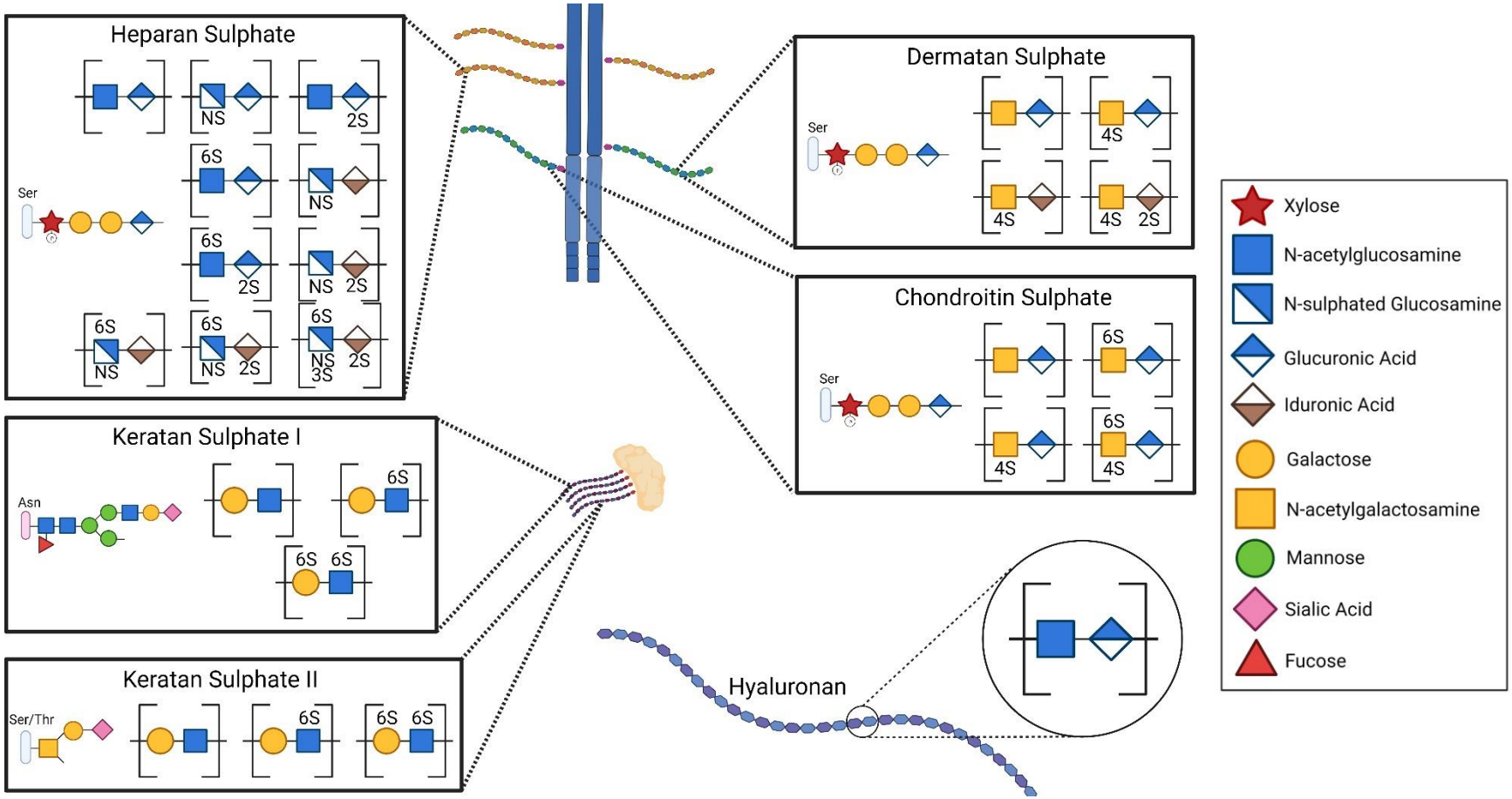


Figure 2 – Glycosaminoglycan Disaccharide Composition. Each GAG is composed of different repeating disaccharides that can be sulphated at different locations. HS is composed of GlcNAc/GlcA, GlcNAc can be N-sulphated, 6-O-sulphated or 3-O-sulphated while GlcA can be epimerized to IdoA and both can be 2-O-sulphated. DS is GalNAc which can be 4-O-sulphated and GlcA which can be epimerized and 2-O-sulphated. CS is composed of GalNAc which can be 4-O- or 6-O-sulphated and GlcA. KS is composed of galactose and GlcNAc which can be 6-O-sulphated. HA is unsulphated GlcA/GlcNAc

1.3 Heparan Sulphate Biosynthesis

1.3.1 Tetrasaccharide Formation

CS/DS and HS/heparin are attached to a common tetrasaccharide linkage region which is synthesized onto a serine residue of the core protein (**Figures 2 and 3**). Initially, a xylosyltransferase uses UDP-xylose to link a xylose (Xyl) molecule onto the serine residue of the protein. There are two isoforms of this enzyme (XYLT1 and XYLT2) which are similar in their activity and optimum conditions, but are differentially expressed in mammalian tissues (Götting, Kuhn and Kleesiek, 2007). Following this, two Gal residues are added sequentially to the growing tetrasaccharide chain by β1-4 and β1-3 galactosyltransferase (β4GALT7 and β3GALT6, respectively). The final addition is a GlcA residue which is catalysed by the glucuronyltransferase enzyme (β3GAT3) (Esko, Kimata and Lindahl, 2009). The overall resulting sequence is GlcA-β1,3-Gal-β1,3-Gal-β1,4-Xyl-β-O-Ser with the serine residue situated in a Ser-Gly/Ala-X-Gly consensus peptide sequence – where X stand for any amino acid (Sugahara and Kitagawa, 2002). Similar to the GAG chain itself, the tetrasaccharide linkage region is subject to modification. The Xyl can be phosphorylated at the C2 position while the Gal residues can be phosphorylated at C4 or C6. However, the Gal modifications are restricted to CS/DS chains whereas the Xyl modification has been shown to be present in HS/heparin as well (Tone *et al.*, 2008).

1.3.2 Initiation and Elongation

HS biosynthesis is initiated by the addition of a GlcNAc residue onto the tetrasaccharide linker (**Figure 3**). This addition acts as the 'switch' to initiate HS biosynthesis and not CD/DS synthesis which is initiated by the addition of a GalNAc residue. The GlcNAc unit is added by exostosin-like glycosyltransferase enzymes, EXTL1, EXTL2 and EXTL3. EXTL3 is the only EXTL enzyme present in *Drosophila,* while vertebrates have been shown to have all three subtypes. However, the role of EXTL1 and EXTL2 in HS biosynthesis is unclear and as *Drosophila* only have EXTL3, it suggests the other two are not essential (Busse *et al.*, 2007). Contrary to this, a small interfering RNA (siRNA) mediated down-regulation of EXTL2 in human embryonic kidney (HEK) 293 cells

results in an increase in HS chain length indicating EXTL2 plays a negative regulatory role in chain polymerization (Katta *et al.*, 2015)

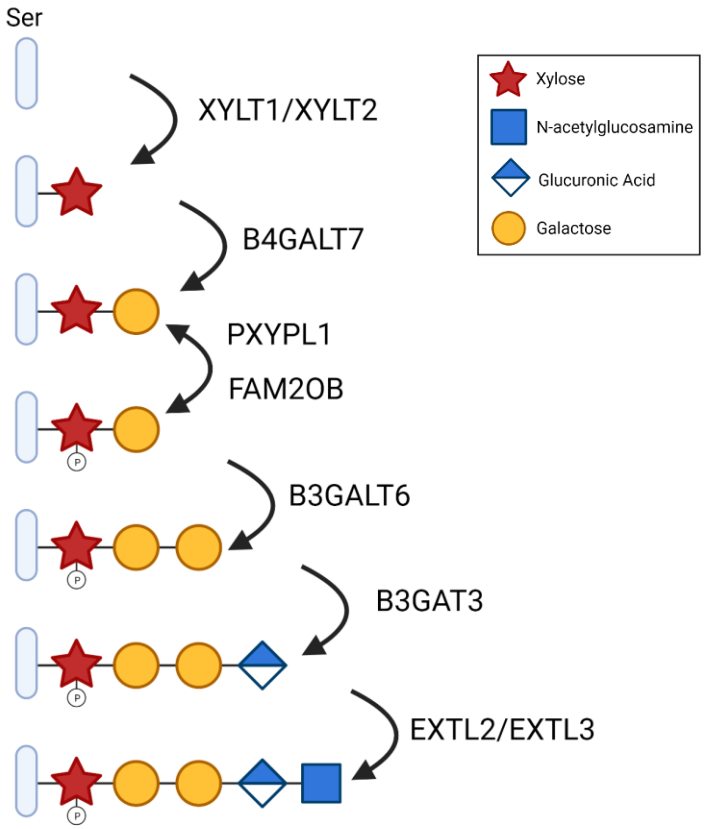


Figure 3 - Tetrasaccharide Linker Synthesis. The tetrasaccharide linker is synthesised onto serine residues of core proteins. Xylose (which can be reversibly phosphorylated) is added first followed by two galactose residues and a glucuronic acid. The linker region is the same for HS/CS/DS. The addition of an N-acetylglucosamine residue by EXTL2/EXTL3 initiates HS biosynthesis.

After initiation, two polymerase enzymes (exostosin glycosyltransferase 1 and 2; EXT1 and EXT2) act on the HS chain (**Figure 4**). They elongate the chain by adding GlcA and GlcNAc residues alternately. The enzymes are thought to be bifunctional, but both are required for correct HS polymerization (McCormick *et al.*, 2000). Recent work, focused on high-resolution cryo-electron microscopy, has uncovered the mechanisms behind this complex assembly. It is thought that formation of the complex is essential for protein stability and the tightly packed complex is held in place by opposing charged residues – EXT1 is more positively charged while EXT2 is more negatively charged. The formation of the

EXT1/EXT2 complex is likely to make HS biosynthesis more efficient (Leisico *et al.*, 2022).

Some of the first mutational studies were carried out in *Drosophila. Tout-velu (ttv)* is a *Drosophila* mutant homologue of *Ext.* The mutant shows segment polarity and a phenotype similar to that of *wingless (*wg; a member of the Wnt family). Unlike other *Drosophila* mutants (discussed in section 1.3.3), the *ttv* mutant displays normal mesoderm migration, indicating FGF signalling is normal. As both wg and Hedgehog (Hh) signalling are implicated in segment polarity, two wg dependent processes were investigated, and appeared the same as wild type (WT). The conclusion drawn was that *ttv* affects Hh signalling. This was further confirmed by the absence of a key protein known as Patched which is usually expressed as a result of Hh signalling (Bellaiche, The and Perrimon, 1998; The, Bellaiche and Perrimon, 1999).

It has been shown that *EXT1* mRNA is present from very early embryogenesis and continues to be present throughout development. It is detectable in all adult mouse tissues, indicating it has an essential role. While $Ext1^{+/-}$ mice appear phenotypically normal, $Ext1^{-/-}$ mice show embryonic lethality and closer investigation reveals that the embryos fail to develop the extraembryonic region and hence fail to gastrulate. At E7.5, the knockout embryos remain as two-layered cylinders of endoderm and ectoderm and by E8.5 the embryos were resorbed. Radiolabelled HS of $Ext1^{-/-}$, $Ext^{+/-}$ and WT mouse embryonic stem cell (mESC) lines revealed that heterozygous KO cells produce 2.3-fold less HS and homozygous KO mESCs produce no HS. An increase in CS production is seen in both of the cell lines lacking Ext1 (Lin *et al.*, 2000). Similarly, loss of Ext2 results in early embryonic lethality due to gastrulation failure while heterozygous mice appear phenotypically normal. Upon close examination of $Ext2^{+/-}$ mice, it was found that they form one or more exostoses (non-malignant cartilage capped tumours) on the ribs. The HS analysis of the *Ext2* knockouts revealed similar results as *Ext1* knockouts – decreased HS in heterozygous lines and no HS in homozygous lines (Stickens *et al.*, 2005).

mESCs with homozygous mutations in *Ext1* were generated and characterized. As expected, the cells produced no HS. It was found that *Ext1*$^{-/-}$ mESCs were able to maintain pluripotency in the absence of HS but were unable to differentiate. The addition of exogenous heparin in the cultures partially restored the differentiation capabilities (Johnson *et al.*, 2007; Kraushaar, Yamaguchi and Wang, 2010; Holley, Pickford, *et al.*, 2011; Pickford *et al.*, 2011). These finding indicate that HS is required for mESCs to exit the pluripotent stage and commit to the formation of different lineages.

More recently, human induced pluripotent stem cell (hiPSC) *EXT1*$^{+/-}$ and *EXT2*$^{-/-}$ lines have been created by CRISPR/Cas9 knockout (Jamie Thompson, personal communication). The genes are disrupted by the insertion of an antibiotic resistance cassette resulting in the heterozygous and homozygous genotypes. These hiPSC lines show similar changes to HS synthesis as that observed in the mouse lines – reduced HS in heterozygous lines and no HS in homozygous lines. Strikingly, the CS/DS population appeared to be of a similar level in KO hiPSCs compared to WT hiPSCs, but the KS population was significantly increased. It was extremely surprising that these cells were able to maintain their pluripotency as HS is required for FGF2 signalling and FGF2 is required for the activation of pluripotency genes (Turnbull *et al.*, 1992; Maccarana, Casu and Lindahl, 1993; Vallier, Alexander and Pedersen, 2005). It is important to note that this is unexpected in hiPSCs as their signalling requirements differ from mESCs which do not rely on FGF2 for pluripotency (Mossahebi-Mohammadi *et al.*, 2020). The knockout hiPSC lines were able to differentiate to all three germ layers which is something KO mESCs were unable to do in the absence of HS (Jamie Thompson, personal communication).

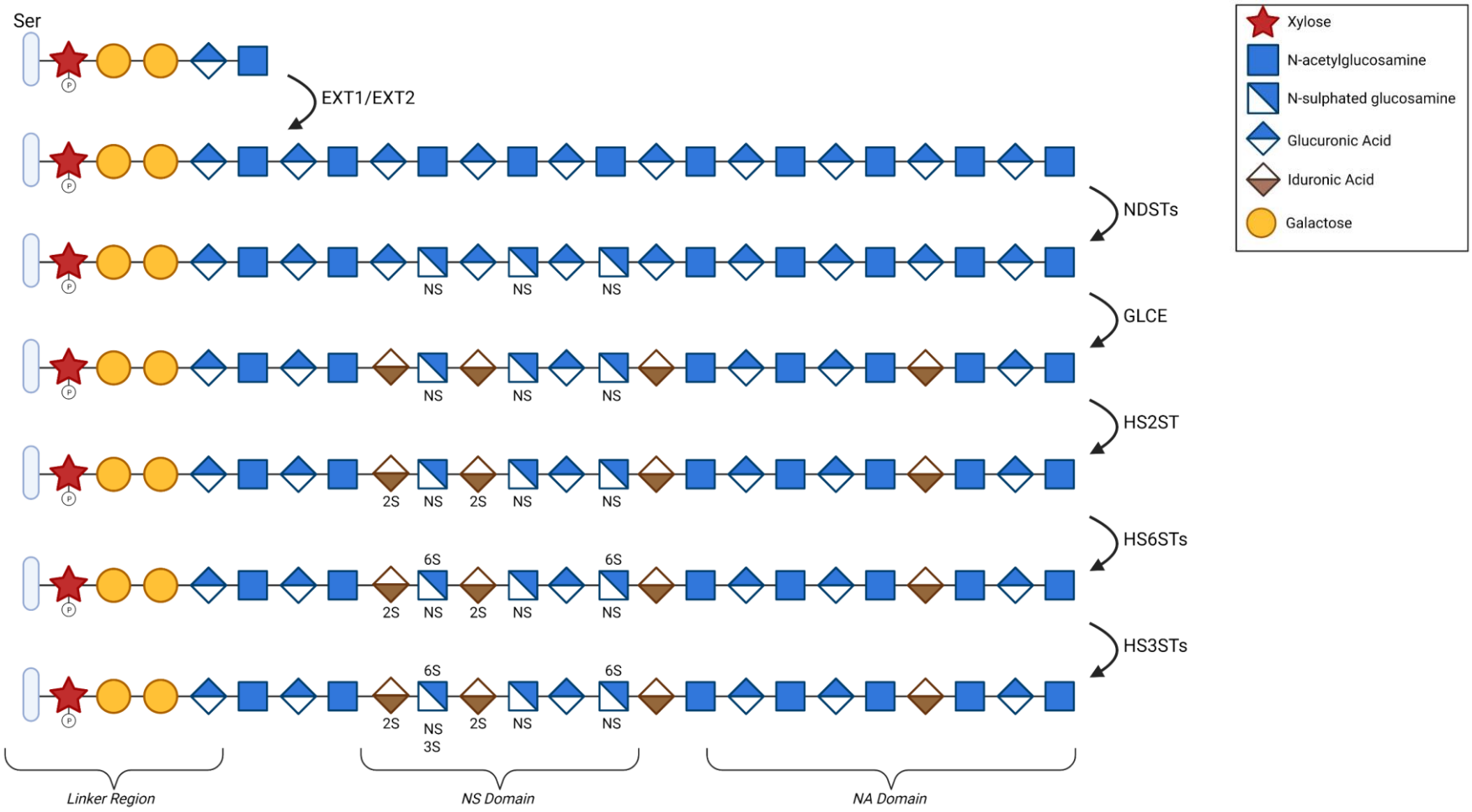


Figure 4 - Heparan Sulphate Biosynthesis. Following linker synthesis, the HS chain is elongated by EXT1/EXT2 which adds repeating GlcNAc/GlcA residues. The chain is then subject to modification which results in heterogenous chains of N-sulphated and N-acetylated domains.

1.3.3 N-Deacetylation and N-sulphation of Glucosamine

Modification of the HS chains occurs while elongation is still taking place. It is important to note that the modifications are all incomplete, yet interdependent. The first committed step in the modification series is N-deacetylation followed by N-sulphation. This is catalysed by a group of bifunctional enzymes known as N-deacetylase/N-sulfotransferases (NDSTs). Acetyl groups are removed from the GlcNAc residues of the HS chain which generate free amino groups. These are then sulphated through the N-sulphotransferase activity of the NDSTs. The NDST family of enzymes converts GlcNAc residues to N-sulphated glucosamine (GlcNS) residues (Kjellén, 2003).

It was initially reported that there were two family members of the NDST family – NDST1 and NDST2. It was hypothesized that NDST1 was involved in the modification of HS and NDST2 was involved in the modification of heparin meaning the latter would only be present in mast cells. However, this theory has been disproved since NDST2 has been shown to be widely expressed, including in tissues and cells that do not produce heparin (Kusche-Gullberg *et al.*, 1998; Toma, Berninsone and Hirschberg, 1998).

Sulphateless (sfl) is a *Drosophila* mutant homologue of *Ndst.* The mutant shows similar segment polarity to the *ttv* mutant. Heartless (Htl) and breathless (Btl) are two mitogen activated protein kinase (MAPK)-activating FGF receptors whose activity is significantly reduced in *sfl* mutants. Reduced mesoderm migration is seen in all three mutants (Lin *et al.*, 1999; The, Bellaiche and Perrimon, 1999). These findings, along with the *ttv* mutant discuss previously, indicate that GAGs, and their subsequent sulphation, are essential for FGF and Wnt signalling, both of which are known to be implicated in embryonic development (Häcker, Lin and Perrimon, 1997).

KO mice have provided much information on the role and relationship of HS biosynthetic enzymes. Two Ndst1 mouse models were created – a heterozygous knockout and a homozygous knockout. The *Ndst1*$^{-/-}$ pups died in the neonatal period – after 5 minutes but before 10 hours. The pups were presumed to have

severe respiratory issues due to their gasping-style breathing and cyanotic skin discolouration. This was confirmed by the autopsy that showed severe developmental defects of the empty lungs. Histological analysis revealed that the lungs of the homozygous KO mice had normal branching, but the lungs were collapsed or closed preventing air from entering (Fan *et al.*, 2000). Furthermore, immunostaining showed that the number of type II epithelial cells within the lungs was increased, yet the differentiation of these cells appeared delayed leading to a reduction in surfactant deposition (Ringvall *et al.*, 2000). The lungs appeared to be the only organ affected by the gene knockout. HS disaccharide analysis on *Ndst1* null embryos showed the number of N-sulphated disaccharides was greatly reduced due to the lack of the Ndst enzyme. In addition to N-sulphation, the 2-O-sulphation modification is reduced considerably along with a slight decrease of the 6-O-sulphation modification (Fan *et al.*, 2000). Less IdoA residues were also noted, indicating a reduction in epimerization of GlcA. These findings highlighted the importance of the role of the NDSTs in regulation of further modification of the HS chain (Forsberg and Kjellén, 2001). No difference between heterozygous pups and WT pups were reported and HS disaccharide analysis from the heterozygous mice revealed only minor changes in the HS composition with small decreases in the N-sulphated residues (Fan *et al.*, 2000). This indicated that when there is total loss of Ndst1, Ndst2 somewhat compensates but N-sulphation cannot be fully maintained (Grobe *et al.*, 2005).

Ndst2 null mice are viable and fertile and display no obvious defects, including in the lungs. Both male and female offspring are fertile. However, when experiments were carried out to determine whether *Ndst2*$^{-/-}$ cells could produce heparin and HS, it was found that cells were producing no heparin, and less HS when compared with WT mice. Furthermore, there was a general decrease in the number of connective tissue-type mast cells and those present had altered biochemical composition and morphology. Although less HS was synthesized in the cells of the *Ndst2* null mouse, it was demonstrated that the levels of N-sulphation remained the same, indicating that the other isoforms were compensating for the lack of isoform 2 (Forsberg *et al.*, 1999). This finding contradicts the outcome of previous studies indicating that Ndst2 is not limited to heparin production in the mast cells and is active more widely.

Therefore, although Ndst2 may not be limited to heparin modification, it is clear that it is the primary isozyme involved in its biosynthesis.

As the *Ndst1*$^{-/-}$ knockouts survived embryogenesis, Ndst2 could be partially compensating for the loss of the other isoform. A model was created that introduced a dual KO of *Ndst1* and *Ndst2*. The KO was lethal, and offspring did not survive past early embryogenesis. An mESC line was created, and the HS analysed. These cells produced no N-sulphated residues in the HS when compared with control cell lines. Strikingly, the *Ndst1*$^{-/-}$*/Ndst2*$^{-/-}$ cells still had around 50% of the 6-O-sulphated residues compared to the controls. This finding was confirmed by PCR which demonstrated 6-O-sulphotransferase expression in the knockout cell line. An mESC line with a total KO of *Ndst2* and heterozygous KO of *Ndst1* was also analysed. This cell line exhibited a decrease in the overall amount of N-sulphation and was comparable to the *Ndst1*$^{-/-}$ KO with a greater reduction of 2-O-sulphation when compared with 6-O-sulphation (Holmborn *et al.*, 2004). These findings support the idea that Ndst2 somewhat compensates for Ndst1 activity when there is complete loss of isoform 1.

NDST studies were further complicated when two more isozymes were reported – these are known and NDST3 and NDST4. Unlike NDST1 and NDST2, NDST3 and NDST4 have differing levels of catalytic activity, with the former having stronger deacetylase activity and the latter having a stronger sulphotransferase activity. These two NDST isoforms differ dramatically from those initially reported. It has been proposed that this is due to substrate specificity. The catalytic activity of the isozymes varies depending on the degree of epimerization or sulphation. For example, NDST4 may have a greater affinity for HS chains that have already undergone sulphation and epimerization (Aikawa *et al.*, 2001). The four NDST isoforms are also differentially expressed; NDST1 and NDST2 are expressed in all embryonic and adult tissues whereas NDST3 and NDST4 are predominantly expressed in embryonic development and in the adult brain, with NDST4 only present at low levels (Grobe *et al.*, 2005; Jao *et al.*, 2016). KO mice of the newer isoforms have also been studied. *Ndst3*$^{-/-}$ mice have normal development and fertile offspring and show only mild behavioural

abnormalities and differences in haematology. Disaccharide analysis shows that the HS produced by these mice is only moderately under-sulphated. Development of mice with a double KO of *Ndst2* and *Ndst3* is normal whereas those with a dual deletion of *Ndst1* and *Ndst3* present with development issues and a severe decrease in HS sulphation patterning (Pallerla *et al.*, 2008). *Ndst4* null mice were even more mildly affected with no changes in viability, fertility or haematology when compared with wild type mice. The only observed differences were minor changes in the cell makeup of the colon, suggesting that *Ndst4* is involved in colon development and homeostasis (Jao *et al.*, 2016).

1.3.4 Epimerization of Glucuronic Acids

Following N-sulphation of specific residues, the next modification is C5 epimerization of some GlcA residues to their C5 epimer, IdoA (Li *et al.*, 2003). Epimerization is catalysed by a single enzyme – HS C5 epimerase (GLCE) (Crawford *et al.*, 2001). The C5 hydrogen of the GlcA is removed and replaced by hydrogen from a water molecule. IdoA units are usually found linked to a GlcNS residue at the non-reducing end and the disaccharide bound at C1 of IdoA can be N-sulphated or N-acetylated. Therefore, the preferred substrate site for C5 epimerase must be GlcNS-GlcA-GlcNS/Ac where the GlcA is epimerized to an IdoA residue (Maccarana *et al.*, 1996). It is important to note that epimerization of GlcA will not occur if the non-reducing end of the substrate is O-sulphated (Jacobsson *et al.*, 1979). This is linked to the hierarchy of HS modification which is related to the order of biosynthesis seen in **Figure 4**. This hierarchical order means that once the chain has moved onto O-sulphation of specific residues, epimerization of those residues can no longer occur. It is important to note that the hierarchical order can be subverted, as discussed above with the *Ndst* knockouts. However, it is generally assumed that this hierarchical order is followed in native HS biosynthesis where machinery has not been removed or altered. The C5 epimerization modification is readily reversible when in solution meaning the C5 epimerase can convert GlcA to IdoA and IdoA to GlcA (Hagner-McWhirter, 2000). The reversibility of the reaction was also studied in cellular systems using a radioactive label and HS structure analysis. It was found that the reaction was reversible and noted that most IdoA residues within N-sulphated domains were 2-O-sulphated which provided a

potential explanation for the resistance of the reverse reaction. Furthermore, IdoA residues found in domains of alternating GlcNAc/GlcNS were rarely 2-O-sulphated but were sulphated at the C6 position, further supporting the idea that O-sulphation renders C5 epimerase inactive (Hagner-McWhirter *et al.*, 2004).

HS C5 epimerase is encoded for on the *GLCE* gene. Heterozygous knockout of the *Glce* gene results in neonatal lethality from respiratory failure in mice. Further investigations were conducted, and it was found that the mice also lack kidneys, have bilateral iris coloboma and several skeletal abnormalities including shorter stature, missing bones and twisted tails. The lungs were deflated and had thicker alveolar walls. All other organs appeared unaffected. Analysis of the HS structure was performed, and it was shown that the IdoA residues were completely absent from the *Glce*$^{-/-}$ mouse, but the heterozygous KO was indistinguishable from the WT. The homozygous KO mouse showed an increase in the GlcA residues, some of which were 2-O-sulphated. The overall number of N-sulphated and O-sulphated residues remained the same across *Glce*$^{-/-}$, *Glce*$^{+/-}$ and *Glce*$^{+/+}$ mice. These findings were striking considering the importance of HS in early development and demonstrated that the IdoA residue may not be necessary for all HS interactions (Li *et al.*, 2003).

1.3.5 2-O-Sulphation

The first of the O-sulphotransferases to act on HS chains is HS 2-O-sulphotransferase (HS2ST1). There is a single 2-O-sulphotransferase enzyme, and its role is to transfer a sulphate group to the C2 position of hexuronic acid. 2-O-sulphation of IdoA residues is a universal component of HS chains with IdoA(2S)-GlcNS forming a key backbone of the HS sulphated domain (Merry *et al.*, 1999). GlcA residues can be 2-O-sulphated, but this modification is more rare as Hs2st1 shows preference for the IdoA residues (Maccarana *et al.*, 1996; Pinhal *et al.*, 2001). Hs2st1 expression is known to be widespread in early development and plays a vital role in FGF signalling (Merry *et al.*, 2001).

The *Hs2st1* KO mouse displayed bilateral renal agenesis, skeletal deformities and eye defects but the remainder of their

development was relatively unaffected. Having said this, the mice were not viable and died at birth (Bullock *et al.*, 1998). Disaccharide analysis of the HS was performed, and it was found that the mice completely lack the 2-O-sulphation modification. This total loss was coupled with an increase in N- and 6-O-sulphation with the 6-O-sulphation confined to the N-sulphated residues, i.e. UA-GlcNS(6S) rather than UA-GlcNAc(6S) (Merry *et al.*, 2001). The 2-O-sulphation modification of IdoA residues has been shown to be highly important in binding FGF2 to the cell surface and is a requirement for its signalling (Turnbull *et al.*, 1992). With this knowledge, protein binding assays were performed with GAGs isolated from the *Hs2st1*$^{-/-}$ mouse. Although binding affinity for FGF1 and FGF2 was decreased in HS from the *Hs2st1*$^{-/-}$ mouse when compared to the WT, cell lines still produced a response to the growth factors. The response to the growth factors was determined using an assay that assesses activation of MAPK. As the *Hs2st1*$^{-/-}$ mouse displayed a complete lack of 2-O-sulphation, it was extremely surprising that IdoA(2S) dependent factors were still eliciting a response. This finding challenged the belief that IdoA(2S) is a requirement for FGF2 binding and signalling. It was concluded that the increase in N- and 6-O-sulphation resulted in the overall charge density and distribution throughout the HS chain being maintained, thereby offsetting the lack of 2-O-sulphation. It has also been suggested that FGF2 signalling does not necessarily require high affinity binding which provides an explanation for the FGF activity response, despite a decrease in binding affinity (Merry *et al.*, 2001). The findings from this study imply that the individual modifications associated with HS biosynthesis are not necessarily independent of one another and that alterations in one modification stage can have profound effects on the activities of enzymes carrying out other modifications.

The *Hs2st1*$^{-/-}$ mouse shows striking similarities to the *Glce*$^{-/-}$ mouse – both have bilateral kidney agenesis, eye and skeletal defects. However, the *Glce*$^{-/-}$ mouse has clear respiratory defects which are likened to that of the *Ndst1* null mouse. The disaccharide analysis from the *Glce*$^{-/-}$ mouse shows increased N-sulphation and increased 6-O-sulphation – similar to the *Hs2st1*$^{-/-}$ mouse. Due to the complete lack of IdoA residues in the *Glce*$^{-/-}$ mice, there was no 2-O-sulphation of IdoA recorded, but there was a significant increase in the 2-O-sulphation of GlcA residues (Li *et al.*, 2003). This finding

aligns with the theory of the enzymes not being completely independent and that alteration in HS modification causes further effects on other enzymes involved in the pathway. A study which involved relocation of modification enzymes to the endoplasmic reticulum (ER) found that relocation of GLCE resulted in redistribution of the 2-O-sulphotransferase. These two enzymes co localize *in vivo,* and in HS2ST1 deficient cells, the C5 epimerase was relocated to the ER, but upon 2-O-sulphotransferase cDNA transfection, the C5 epimerase relocated back to the Golgi. This study also found that C5 epimerase activity was decreased in the mutant cell line, but when 2-O-sulphotransferase activity was restored, epimerase activity increased. These finding suggest that these two enzymes are somewhat dependent on each other, and both are required for full activity (Pinhal *et al.*, 2001)

1.3.5.1 Heparan Sulphate 2-O-Sulphation and Kidney Development

Kidney formation is the result of epithelial-mesenchymal transition, ureteric bud outgrowth and metanephric mesenchyme condensation. An HSPG-rich matrix lies between the two developing tissues and acts as a potential channel for growth factors and signalling molecules. Additionally, Hs2st1 is expressed in both the ureteric bud and metanephric mesenchyme of developing mouse kidneys so the requirement for 2-O-sulphated HS could occur at several distinct locations (Merry and Wilson, 2002).

At 10.75 days post conception in mice, the ureteric bud emerges from the Wolffian duct (the earliest epithelial component of the urogenital system) and comes into contact with the metanephric mesenchyme. This interaction causes the ureteric bud to elongate and bifurcate while the metanephric mesenchyme condenses into epithelium and differentiates (Bullock *et al.*, 1998; Merry and Wilson, 2002; Faa *et al.*, 2012). When kidneys of *Hs2st1*$^{-/-}$ mice were compared against those of wild type litter mates, it was found that at 11.5 days post conception the wild type kidneys had bifurcated once and the metanephric mesenchyme had condensed – something which had not occurred in the *Hs2st1*$^{-/-}$ mice. Furthermore, at 12.5 days post conception, condensation and bifurcation was still not seen in knockout mice while wild type kidneys had continued to develop – indicating that kidney

development was not delayed but completely halted (Bullock *et al.*, 1998). Ureteric bud outgrowth from the Wolffian duct still occurred in knockout mice indicating that Hs2st1 is not required at this stage of development, but rather for the metanephric mesenchyme condensation and subsequent ureteric bud bifurcation. In humans, a similar pathway is observed, with the ureteric bud outgrowth beginning at day 28 post conception (Faa *et al.*, 2012).

Various signalling factors are required for correct development of the kidneys including glial cell line-derived neurotrophic factor (GDNF) which is secreted by the undifferentiated mesenchymal cells and interacts with c-ret tyrosine kinase receptor found on the ureteric bud. Correct signalling of GDNF, and its control by PAX2, is considered essential for proper ureteric bud outgrowth and branching. Mice lacking c-ret show striking similarities in phenotype when compared with *Hs2st1* knockout mice and die perinatally due to bilateral renal agenesis. Upon closer investigation, it appeared kidney development was halted at various stages with the most frequent being at 11.5 days post conception, due to failure of ureteric bud outgrowth (Schuchardt *et al.*, 1994). This indicates that GDNF does not interact with Hs2st1 for ureteric bud outgrowth, but an interaction could still be required for bifurcation.

Another key signalling pathway involved in kidney formation is FGF signalling. It is well known that FGF2 interacts with the IdoA(2S) motif of HS to potentiate high affinity binding, which provides a tempting explanation to altered kidney development of *Hs2st1*$^{-/-}$ mice (Turnbull *et al.*, 1992; Maccarana, Casu and Lindahl, 1993). When the binding affinities of WT and *HS2ST1*$^{-/-}$ HS to various proteins was assessed, it was discovered that FGF2 bound more weakly when compared to ligands that were less dependent on 2-O-sulphation containing sequences within HS for binding such as hepatocyte growth factor/scatter factor (HGF/SF) and fibronectin. However, despite the reduced affinity, FGF2 was still able to elicit a response – measured by extracellular signal-regulated kinase (ERK) phosphorylation (Merry *et al.*, 2001). It has also been reported that at least one GlcNS(6S) residue is required for FGF2 signalling and as the *Hs2st1*$^{-/-}$ HS has significantly more 6-O-sulphation, it is possible that the overall charge density of the HS is maintained and is still able to elicit downstream signalling (Pye *et*

al., 1998; Merry *et al.*, 2001). However, more precise and high affinity signalling may be necessary for the initial stages of kidney formation.

1.3.6 6-O-Sulphation

The next modification enzyme to act on the HS chain is HS 6-O-sulfotransferase (HS6ST). The enzyme was first isolated in Chinese hamster ovary (CHO) cells. When the isolated enzyme was combined with GlcNS, a sulphate group was added at the C6 position, but there were no modifications at the C2 position of iduronic acid (Kobayashi *et al.*, 1996). Following its isolation, it was transfected into other cells, and it was reported that Hs6st activity was upregulated with no changes in Hs2st1. This finding indicated that the two HS modifications were carried out by different enzymes (Habuchi, Kobayashi and Kimata, 1998).

Three isoforms of HS6ST have been reported - HS6ST1, HS6ST2 and HS6ST3 and each was reported to have differential expression and substrate specificity. In the initial study that isolated the three different isoforms, it was found that HS6ST1 was at high concentration in the liver while HS6ST2 was expressed in the brain and spleen (Habuchi *et al.*, 2000). HS6ST3 was reported to be ubiquitously expressed. Furthermore, it was found that each of the isoforms had different substrate preferences and sulphation activity. HS6ST1 shows preference for IdoA-GlcNS while HS6ST2 has a marked preference for GlcA-GlcNS. The third isoform showed no substrate preference. This study also concluded that HS6ST1 activity was concentration dependent whereas the activity of the other HS6STs was not. The substrate preference aligns with the location of each of the enzymes as IdoA(2S)-GlcNS is at high levels in the liver while GlcA-GlcNS is present at high levels in the brain. These results led to the suggestion that HS6ST1 was involved in sulphation of GlcNAc residues to form N-sulphated domains (Habuchi *et al.*, 2000). This study did not highlight an isoform or mechanism for the 6-O-sulphation of GlcNAc residues. This could be due to the fact that it was conducted *in vitro*, as an *in vivo* study the following year revealed HS6ST1 can add a 6-O-sulphation modification onto GlcNAc residues (Zhang, Beeler, *et al.*, 2001).

Since the original discovery of each of the 6-O-sulphotransferase isoforms and the investigation into their substrate preference, further studies have been carried out. In 2003 a study concluded that all three isozymes catalysed the 6-O-sulphation of GlcA-GlcNS and IdoA-GlcNS residues where the IdoA can be 2-O-sulphated or not. However, HS6ST1 showed a preference for substrates that were not 2-O-sulphated. Furthermore, 6-O-sulphation of these residues is preferred at lower concentration of enzyme (Smeds *et al.*, 2003). A second study identified HS6ST2 and HS6ST3 as having a preference for 2-O-sulphated substrates (IdoA(2S) containing species), with the specificity being three to five times higher when there are two or three 2-O-sulphate groups on the substrate (Jemth *et al.*, 2003).

Activity studies of 6-O-sulphotransferase are much easier in simpler organisms that only have one isoform. *Drosophila* for example has only a single 6-O-sulphotransferase orthologue which is found in cells that express certain FGF receptors. Disruption of the gene encoding the modification enzyme results in reduction of FGF-MAPK activation and a phenotype similar to that seen in FGF mutants (Kamimura *et al.*, 2001). Another simple organism, *Caenorhabditis elegans*, was also found to only have one 6-O-sulphotransferase isoform (Hst-6). In contrast to *Drosophila*, the worms with a 6-O-sulphotransferase knockout are viable and fertile. This indicates that Hst-6 has a less fundamental role, but it has been shown to be involved in neural development (Bülow and Hobert, 2004). In zebrafish, knockout of the enzyme responsible for 6-O-sulphation of HS results in severe muscle degeneration due to impairment of muscle differentiation and somite specification (Bink *et al.*, 2003). Due to the simplicity of these organisms, they are not representative of the phenotype that may been seen in more complex organisms such as mice or humans.

Since the earlier studies in simple organisms, knockouts of the different 6-O-sulphotransferase enzymes in mice has been carried out. *Hs6st1*$^{-/-}$ mice show embryonic lethality, with death occurring in the later stage; embryos appeared viable in the early stages of development (Habuchi *et al.*, 2007). The mice showed severe growth retardation and no hind limb phalange and either smaller,

or no, hind limb tarsus. Analysis of the HS from the KO mice showed a large decrease in the 6-O-sulphation of kidney and lung tissues where HS6ST1 is usually at high levels. 6-O-sulphation was slightly decreased in tissues where Hs6st2 is expressed at high levels – the brain and spleen. UA-GlcNS(6S) and UA-GlcNAc(6S) were reduced further compared with IdoA(2S)-GlcNS(6S) which actually showed an increase in spleen tissue. This increase could be due to the action of HS6ST2 which showed a preference towards 2-O-sulphated residues (Habuchi *et al.*, 2007). In contrast to *Hs6st1* null mice, *Hs6st2* knockouts were viable and fertile. HS analysis of *Hs6st2* null mice displayed a significant decrease of 6-O-sulphated residues, including UA-GlcNS(6S) and UA-GlcNAc(6S), but showed only a slight reduction of UA(2S)-GlcNS(6S) residues. *Hs6st1*$^{-/-}$*/Hs6st2*$^{-/-}$ dual knockout mice are non-viable and death occurs around 14.5 to 15 days post conception (Anower-E-Khuda *et al.*, 2013). HS analysis of *Hs6st1*$^{-/-}$*/Hs6st2*$^{-/-}$ mouse embryonic fibroblasts (MEFs) showed all 6-O-sulphate containing disaccharides to be barely detectable. This indicated that when only *Hs6st1* or *Hs6st2* are knocked out, the other isoform compensates for the loss, whereas when both are null, *Hs6st3* cannot compensate. In both *Hs6st2* null mice and the dual knockouts, 2-O-sulphation was significantly increased (Sugaya *et al.*, 2008).

1.3.7 3-O-Sulphation

The final sulphotransferase to act on the HS chains is the rarest, HS 3-O-sulphotransferase (HS3ST) catalyses the addition of a sulphate at the C3 position of glucosamine residues. The modification is so rare that often it is completely absent from HS chains, and when present it is only in limited numbers. As mentioned earlier, the modifications are interdependent meaning that previous modification steps create the substrate needed for the next modification enzyme. This means that sulphation occurs in discrete domains and that 3-O-sulphation only occurs on previously sulphated residues. The family of 3-O-sulphotransferases is the largest of all the sulphotransferase enzymes with seven family members in vertebrates – HS3ST1, 2, 3a, 3b, 4, 5, 6. They can be categorized into two subgroups based on the structural similarity of their sulphotransferase domain. HS3ST1 and 5 form one group while the others form the second group. Each of the HS3STs has preferential binding sites to the HS chain (Thacker *et al.*, 2014).

HS3ST1 has preference for GlcNS(6S) residues next to a GlcA residue which is not 2-O-sulphated. Furthermore, it will act on glucosamine next to IdoA residues, but the 2-O-sulphation inhibits its catalytic activity (Nguyen *et al.*, 2012). In contrast, HS3ST2, 3a, 3b, 4 and 6 preferentially act on glucosamine residues situated next to 2-O-sulphated IdoA residues (Zhang, Lawrence, *et al.*, 2001; Wu *et al.*, 2004; Thacker *et al.*, 2014). HS3ST5 has less demanding substrate specificities (Chen *et al.*, 2003).

When the 3-O-sulphation modification was described, it was hypothesized that it was essential for the interaction of HS and heparin with antithrombin (Lindahl *et al.*, 1980). Later it was shown that HS with anticoagulant activity (HS^{act}) had a much higher abundance of 3-O-sulphated disaccharides when compared with non-anticoagulant HS (HS^{inact}), further strengthening the previous theory (Shworak *et al.*, 1996). Having said this, only some of the HS3ST isoforms are responsible for HS^{act} synthesis – namely HS3ST1 and HS3ST5 (Liu *et al.*, 1999; Lawrence *et al.*, 2007). The other group of HS3STs create a binding site for the herpes simplex virus (HSV) glycoprotein D (gD)which allows viral entry into the cells. HS3ST5 has the capacity to create binding sites for both antithrombin and gD (Chen *et al.*, 2003; Thacker *et al.*, 2014).

As well as specific targets and functions, the HS3STs also have differential expression throughout the body. HS3ST1 is widely expressed but is found at highest levels in the kidneys and brain. HS3ST2 is found primarily in the brain, but is also at lower levels in the placenta, heart, lung and skeletal muscles. Similar to HS3ST1, HS3ST3 is expressed in most tissues, but the two isoforms – a and b – are found at different levels and different ratios. For example, HS3ST3a is expressed at high levels in the heart and placenta while HS3ST3b is expressed at high levels in the liver and placenta. HS3ST4 is the most selective isoform, only being found in the brain (Shworak *et al.*, 1999). HS3ST5 is present in the tissues of the brain, spinal cord and skeletal muscle while HS3ST6 is expressed mainly in the kidney and liver, but is also found in the heart, brain, lungs and testis (Xu *et al.*, 2005).

Due to the key roles of HS3ST1 in anticoagulation, it would be expected that *HS3ST1*$^{-/-}$ mice have a procoagulant phenotype. However, mice were viable and despite reduced anti-Xa activity, a procoagulation phenotype was not seen. Furthermore, unexpected phenotypes were seen including reduced fertility, intrauterine growth retardations and spontaneous eye degeneration (Shworak *et al.*, 2002; HajMohammadi *et al.*, 2003). This is considerably more striking knowing that knockout of the antithrombin gene (*AtIII*) results in embryonic lethality in mice due to subcutaneous haemorrhage. *AtIII*$^{+/-}$ mice appear phenotypically normal and exhibit no differences with blood coagulation (Ishiguro *et al.*, 2000).

1.3.8 Post Synthetic Modifications

Following the sulphation of HS chains by the different O-sulphotransferase enzymes, two endo-6-sulphatases (Sulf1 and Sulf2) catalyse the release of 6-O-sulphate groups from N-sulphated disaccharides within the HS chain (Dhoot *et al.*, 2001; Morimoto-Tomita *et al.*, 2002). The Sulf enzymes are associated with the cell membrane and remove the 6-O-sulphate modification from HSPGs on the cell surface thereby regulating the activity of essential HS ligands and cell processes (Frese *et al.*, 2009).

Mice deficient for Sulf1 or Sulf2 initially appear to have a normal birthweight. However, upon close inspection, *Sulf2* knockout mice have a slightly decreased body weight and smaller litter sizes (Kalus *et al.*, 2009). Mice who have a double KO for both of the Sulf enzymes display significant difference compared to single knockout and wild type phenotypes. Dual KO mice are of extremely low birth weight, have increased mortality and present with a variety of renal and skeletal defects (Holst *et al.*, 2007).

HS disaccharide composition of MEFs was carried out. Cells deficient for both Sulf enzymes display an 80% increase in mono-sulphated disaccharides while *Sulf1* and (independently) *Sulf2* KO cell lines show a 10% and 30% increase respectively. In addition, *Sulf2*$^{-/-}$ MEFs display a decrease in N-sulphated and 2-O-sulphated HS but double KOs show an increase in these two modifications. *Sulf1* knockouts show no significant difference in N-sulphation or 2-

O-sulphation. It was concluded that a functional interaction and influence between the Sulf enzymes and other HS modification enzymes must exist and a feedback loop system was hypothesized (Lamanna *et al.*, 2008). Interestingly, Sulf2 has a CS/DS chain covalently bound to a hydrophilic domain that recognises the HS undergoing modification. This makes Sulf2 a PG and the CD/DS chain location within the enzyme indicates it plays a role in modulating Sulf2 activity and HS recognition (El Masri *et al.*, 2022).

1.4 Heparan Sulphate Biosynthesis and Disease

In humans, defects in the GAG biosynthesis pathways caused by genetic mutations can have severe implications and lead to a range of disorders. This includes defective tetrasaccharide formation which can give rise to Ehlers-Danlos syndromes (EDS) (Mizumoto and Yamada, 2021). EDS are a group of diseases that present with a range of symptoms and severities. All EDS patients display joint hypermobility, stretchy skin and delicate connective tissue. EDS can be caused by abnormalities in multiple ECM molecules, but a subset has been linked to GAG biosynthesis (Syx *et al.*, 2022). One study found that skin fibroblasts from a child with EDS were unable to synthesize enough GAG chains onto decorin, resulting in half the amount of 'mature' decorin PG being synthesized (Quentin *et al.*, 1990). This was later attributed to defects in *β4GALT7* which is responsible for the addition of the first Gal in tetrasaccharide synthesis (**Figure 3**) (Quentin *et al.*, 1990; Okajima *et al.*, 1999). In addition, mutations in *β3GALT6* also result in a similar phenotype and the two genetic mutations have been classified as spondylodysplasia EDS. This group of diseases is characterized by craniofacial dysmorphisms including a short round face, bulging eyes and a narrow mouth, as well as the common EDS symptoms described above (Syx *et al.*, 2022).

As each gene encodes a specific enzyme along the HS biosynthetic pathway, mutations in different genes result in abnormally synthesized HS – it may result in a shorter or differentially sulphated product. As the sulphation pattern is essential for its role, alterations in the biosynthesis can cause developmental defects. Not every enzyme in the HS biosynthesis pathway has an associated disease. For example, mutations in *GLCE* have not been

reported in humans and are presumed to be fatal. Each of the biosynthetic enzymes that is associated with a disease is introduced below, relative to the order of biosynthesis (**Figure 4**).

Hereditary Multiple Exostoses (HME), also known as Multiple Osteochondroma (MO) is a paediatric disease associated with mutations in the *EXT1* and *EXT2* genes. The disease is inherited in an autosomal dominant manner and patients are always heterozygous for mutations – indicating that a homozygous mutation results in failure of gastrulation, as with the mice (Beltrami *et al.*, 2016). The disease is characterized by the presence of osteochondromas – a cartilage capped projection of bone off the existing bone. Complications can arise where patients develop bowed limbs or altered stature as well as the potential for the growths to become malignant (Bovée, 2008). Mutations in the gene result in loss of function, and as a result, abnormal HS. The HS is still present, but levels are reduced, and chains are significantly shorter. Analysis of HS from osteochondromas of patients also revealed the HS to be more sulphated than the HS of healthy patients. This is due to the complex that forms between EXT2 and NDST1 in the absence of EXT1 (Veraldi *et al.*, 2018).

Intellectual disability (ID) encompasses a large group of untreatable neurodevelopmental disorders that can be characterized by cognitive deterioration and behavioural issues. ID is common in consanguineous families and is inherited in an autosomal recessive (AR) manner. A number of genes responsible for ARID have been identified, with mutations in *NDST1* thought to be a significant contributor. There have been 9 specific cases identified, to date, where ARID has been a result of mutations in *NDST1*. In each case, patients were born heavier than average to consanguineous, healthy parents. Children display developmental delays before one year of age with first words and steps generally occurring around three. All of the patients show aggressive behaviour which often leads to self-harm as well as sleep disturbances, communication issues, a lack of understanding of simple commands and a low IQ score. Occasionally, patients present with muscle hypotonia, ataxia and seizures (Reuter *et al.*, 2014; Armstrong *et al.*, 2017).

When this project began, there were no known human diseases associated with mutations in *HS2ST1*. This aligned with the theory that humans could not survive with genetic mutations due to the embryonic lethality displayed in mice. However, during this PhD four individuals, from three families, were found to have alterations in the *HS2ST1* gene (Schneeberger *et al.*, 2020). Individual 1 was five years of age at diagnosis and presented with severe developmental delay, delayed speech and various skeletal abnormalities including short stature, brachydactyly (short fingers and toes) and delayed bone ossification. Patient 2 (14 years old) who is from a different family, also had delayed speech but only moderate developmental delay. As with patient 1, the skeletal abnormalities of patient 2 were varied, and included knock knees, clubfoot, broad thumbs and toes and camptodactyly (bent fingers). Patient 2 also displayed unilateral renal agenesis with an enlarged ureter. Patients 3 and 4 are from the same family – patient 4 was five days old at the time of data collection and patient 3 was a pregnancy terminated at 23 + 2 weeks of gestation. Data collection of neurological features was not possible with these two patients. However, the skeletal abnormalities are similar between the two – flexion contractures of the limbs, in-turned feet, brachydactyly and broad distal phalanges of the hands and feet. Similar to patient 2, patient 4 also has unilateral renal agenesis. Patient 3 has bilateral renal agenesis, as seen in mice, and also a rudimentary bladder. As all of these patients presented with generalised symptoms, a full genome scan was performed which revealed the specific mutations within the *HS2ST1* gene (**Table 1, Figure 5**).

Table 1 - HS2ST1 Patient Mutations

Patient	Mutation Type	Exon	Mutation
1	Compound heterozygous	2 & 4	c.342delA (p.Val115*) c.527T>C (p.Phe176Ser)
2	Compound heterozygous	1 & 4	c.59_62delinsGAA (p.Phe20*) c.493G>T (p.Asp165Tyr)
3	Homozygous	4	c.567A>C (p.Arg189Ser)
4	Homozygous	4	c.567A>C (p.Arg189Ser)

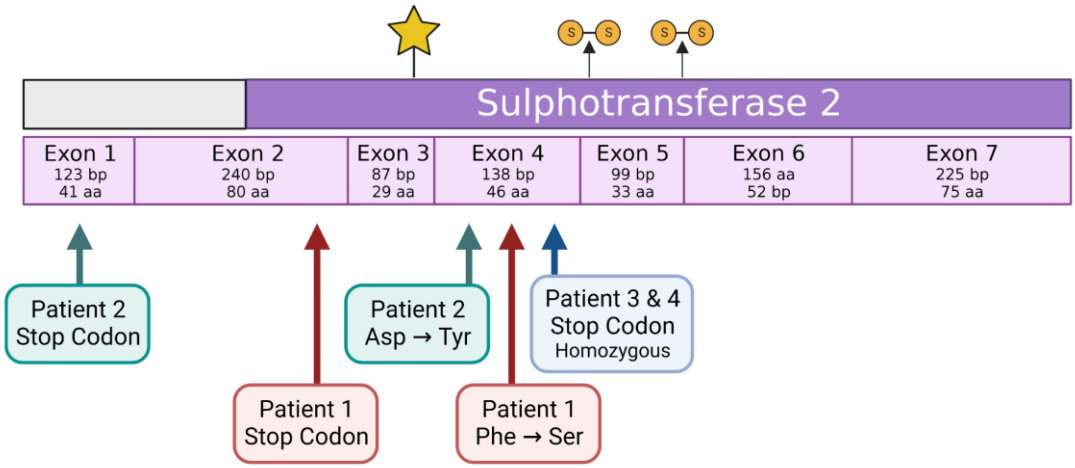


Figure 5 - HS2ST1 Patient Mutations. The HS2ST1 gene is shown with the active site (yellow star), disulphide bonds (yellow circle) which are important for the stability and conformation, exon number, base pair (bp) number and amino acid (aa) number. Patient 1 and patient 2 have a compound heterozygous mutation that results in one stop codon and one amino acid substitution each. Patients 3 and 4 have the same homozygous mutation in exon 4 which results in a premature stop codon.

It has been reported that *HS6ST1* is mutated in patients with a genetically heterogeneous condition known as idiopathic hypogonadotropic hypogonadism (IHH). IHH is characterized by a lack of sexual maturation and infertility. Some patients have an impaired sense of smell, and others have a combination of IHH and insomnia – known as Kallmann syndrome. A paper by Tornbeg *et al.* (2011) reported 7 subjects with mutations in *HS6ST1* and showed that each of the mutations resulted in amino acid substitution – six out of seven of which were heterozygous. It is important to note that several other genes are implicated in IHH and the disease state can occur due to mutations in at least eight other genes (Tornberg *et al.*, 2011).

Although mutations in the *HS3ST* genes are not associated with diseases in the same way as the other biosynthetic enzymes, they have been implicated in several diseases including cancer, Alzheimer's and hereditary angioedema. As mentioned previously, HS3ST2 is found primarily in the brain, more specifically regions susceptible to alterations in the tau proteins during the progression of Alzheimer's disease, characterized by hyperphosphorylation of tau protein. Tau phosphorylation is downregulated when *hs3st2* expression is inhibited in zebrafish. It has also been shown that during the early stages of disease development, HS appears to

accumulate inside the cells, much earlier than tau protein alterations can be detected (Sepulveda-Diaz *et al.*, 2015). In addition to this, *Hs3st1*$^{-/-}$ mouse lung endothelial cell lines show reduced cell surface binding and internalization of the tau protein. Along with the finding that HS3STs are highly upregulated in the disease state, it can be said that the 3-O-sulphation motif plays a key role in the onset and progression of Alzheimer's disease (Zhao *et al.*, 2020). HS3ST2 has also been investigated for its roles in cancer progression. It was initially found that the gene was hypermethylated at the 5' region in tumour tissues, but not in tissues surrounding the tumours, leading to reduced expression (Miyamoto *et al.*, 2003; Denys and Allain, 2019; Gulberti *et al.*, 2020). HS3ST6 has been found to be linked with a novel hereditary angioedema (HAE). HAE is an autosomal dominant disease characterised by recurring oedemas which are localized to various organs. The symptoms include swelling of the skin and tongue, abdominal pain and laryngeal attacks. Whole exome sequencing of three patients, from one multigenerational family, was carried out and revealed a heterozygous missense mutation (c.430A>T) in *HS3ST6* which results in an amino acid substitution (Bork *et al.*, 2021).

1.5 Heparan Sulphate Degradation

HS degradation is initiated at the cell surface and is concluded in the lysosome. The internalization mechanism is unknown, but the endocytic pathway is well documented. Similar to the biosynthesis of HS, degradation occurs in a well-ordered stepwise manner (Freeze, Kinoshita and Schnaar, 2017). However, an exact sequence of enzymes cannot be deduced due to the heterogenous nature of HS chains (Kresse and Glössl, 1987). The first enzyme to act on the HS chain is heparanase 1 (Hep1), an endoglucuronidase that breaks the chain up into oligosaccharides of 10-20 units (Freeze, Kinoshita and Schnaar, 2017). This enzyme shows preference for highly sulphated domains and acts on the GlcNS(6S)-GlcA-GlcNS regions, cleaving the β1-4 GlcA linkage. Hep1 expression is tightly regulated by a myriad of factors including chemokines, growth factors and metabolites. Hep1 exists as a proenzyme which is then cleaved into its active form and released from the cells in response to stimuli. Prior to its secretion, Hep1 is found in the late endosomes and lysosomes and so acts in

an acidic environment. The optimum pH for Hep1 activity is pH 5.5 – pH 6.0 and it shows no activity below pH 3.5 or above pH 7.0, indicating the environmental pH can tightly regulate its activity (Vreys and David, 2007; Mayfosh, Nguyen and Hulett, 2021). Heparanase 2 (Hep2), a protein that shows 35% sequence identity to Hep1 was discovered but has been shown not to have any enzymatic activity with respect to HS degradation. Though there is little information regarding their interaction, it is thought that Hep2 inhibits the activity of Hep1 (Linker and Hovingh, 1968). It is important to note that the heparanase here differs from the heparinase discussed later in the thesis. Three inducible heparinase enzymes have been harnessed from *Flavobacterium heparinum* for use within a laboratory setting (Galliher *et al.*, 1981).

The subsequent sequential degradation of the shortened HS chains occurs in the lysosome. Each of the enzymes is targeted to the lysosome via the mannose-6-phosphate receptor that recognises phosphorylated mannose residues on the enzymes (Linker and Hovingh, 1968). The sequence described is an example of how HS could be broken down but will vary depending on the structure of the HS chain subject to degradation. The following description aligns with the degradation example seen in **Figure 6**. The first of the enzymes to act in the lysosome is iduronate 2-sulphatase (IDS) which de-sulphates the non-reducing terminal IdoA(2S) residues. This modifies the terminal residue and creates the substrate for the next enzyme (Freeze, Kinoshita and Schnaar, 2017). After the activity of IDS, the chain is the correct substrate for α-L-iduronase (IDUA) which breaks the α-L-IdoA glycosidic bond to release the IdoA residue (Linker and Hovingh, 1968).

The terminal disaccharide is now a sulphated glucosamine. These residues must be removed in three distinct steps; sulphate removal, acetylation and cleavage (Freeze, Kinoshita and Schnaar, 2017). N-sulphoglucosamine sulphohydrolase (SGSH), often referred to as sulphamidase, acts on GlcNS residues to remove the sulphate from the amino group. SGSH is different to other sulphatatses involved in HS degradation in that it is an N-sulphatase rather than an O-sulphatase meaning the active site has obvious differences, although the rest of the structure shows similarities. Heparan-α-glucosamide N-acetyltransferase (HGSNAT)

acetylates glucosamine residues that have been desulphated by SGSH. Unlike the other enzymes in the degradation process, HGSNAT is not directly involved in the breakdown of HS and instead modifies the chain to create a substrate which is specific to the next enzyme in the sequence (Linker and Hovingh, 1968). Acetyl-CoA acts as a donor and provides the acetyl group to the N-acetyl transferase (Freeze, Kinoshita and Schnaar, 2017). The resulting GlcNAc residue is then subject to removal by α-N-acetylglucosaminidase (NAGLU) which cleaves the α-GlcNAc glycosidic bond to release the GlcNAc.

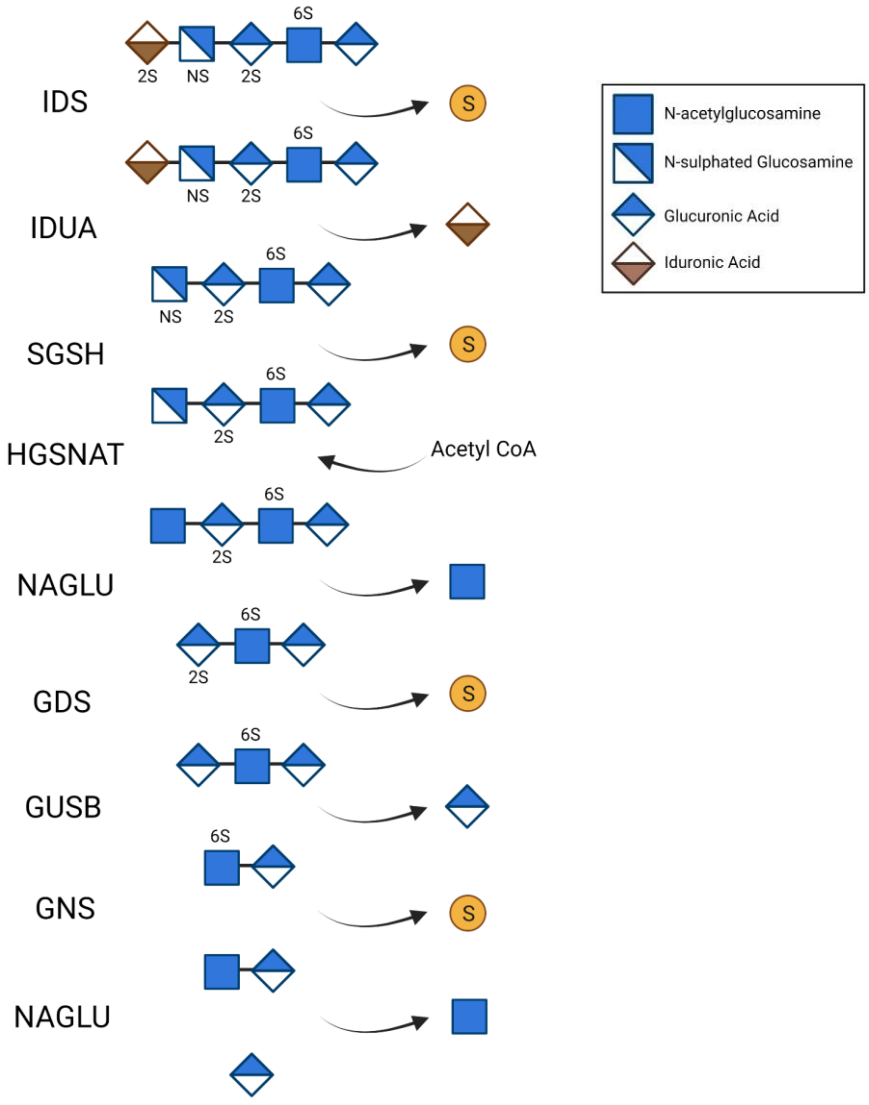


Figure 6 - Heparan Sulphate Degradation. Heparanase 1 first degrades the HS chain into smaller fragments. Each monosaccharide, or sulphate group is reduced from the non-reducing end sequentially. HS degradation also includes a synthetic step in which glucosamine residues must be N-acetylated following N-sulphate removal before the monosaccharide can be cleaved. Defects in this pathway can lead to mucopolysaccharidosis.

The example in **Figure 6** then shows the chain to have a GlcA(2S) residue in the terminal position. Glucuronate 2-sulphatase (GDS) removes the 2-O-sulphate group before subsequent cleavage by β-glucuronidase (GUSB) (Freeze, Kinoshita and Schnaar, 2017). If cleavage of the glucuronic acid reveals GlcNAc(6S) then the sulphate group is removed by N-acetylglucosamine-6-sulphatase (GNS) before subsequent cleavage of the GlcNAc residue by NAGLU (Freeze, Kinoshita and Schnaar, 2017).

Although not shown in **Figure 6**, 3-O-sulphated residues are removed through the action of N-sulphoglucosamine-3-sulphatase, but due to the rarity of this modification, naturally, the degrative enzyme is equally as rare (Linker and Hovingh, 1968).

1.5.1 The Mucopolysaccharidoses

Defects in the HS degradative enzymes result in a group of rare genetic disorders known as the mucopolysaccharidoses (MPSs), a type of lysosomal storage disorder (LSDs). Each different MPS is caused by a defect in a different enzyme in one of the GAG catabolic pathways. Generally, the enzyme alteration results in a truncated form of partially degraded GAG, which are often highly sulphated, accumulating in the lysosomes (Holley, Deligny, *et al.*, 2011; Wilkinson *et al.*, 2012). Abnormal GAGs then begin to accumulate in other areas of the cell and the ECM. The accumulation, coupled with the lack of GAG turnover, affects key cellular function. Although the symptoms, time of onset and prognosis are different for each type of MPS, the continued accumulation of partially degraded material results in cellular damage, leading to cognitive delay, impairment and regression, organ failure and reduced life expectancy (Muenzer, 2011). All of the MPSs, except MPS II, are inherited in an autosomal recessive manner and affect both sexes equally. MPS II is X-linked recessive, and so generally affects males (McKusick, 1965). **Table 2** details each of the different MPSs and their associated enzyme deficiency. Each different type of MPS III and MPS IV is caused by a different catabolic enzyme whereas the subtypes of MPS I are characterised by their severity.

Table 2 - The Mucopolysaccharidoses

Type	Subtype	GAG	Deficient Enzyme
MPS I	Hurler	DS, HS	α-L-iduronidase (IDUA)
	Hurler-Scheie		
	Scheie		
MPS II	Hunter	DS, HS	Iduronate-2-sulphatase (IDS)
MPS III	Sanfilippo A	HS	N-sulphoglucosamine sulphohydrolase (SGSH)
	Sanfilippo B		Alpha-N-acetylglucosaminidase (NAGLU)
	Sanfilippo C		N-acetyltransferase (HGSNAT)
	Sanfilippo D		N-acetylglucosamine-6-sulfatase (GNS)
MPS IV	Morquio A	CS, KS	N-acetylgalactosamine-6-sulphtatase (GALNS)
	Morquio B	KS	B-galactosidase (GLB1)
MPS V	Formerly Scheie Syndrome		
MPS VI	Maroteaux-Lamy	CS, DS	Arylsulphatase B (ARSB)
MPS VI	Sly Syndrome	CS, DS, HS	B-glucuronidase (GUSB)
MPSIX	Natowicz disease	HA	Hyaluronidase (HYAL1)

1.5.2 Mucopolysaccharidosis Type IIIA

MPS IIIA, also known as Sanfilippo A, is caused by a mutation in the *SGSH* gene. SGSH removes the sulphate group from the amino group of glucosamine residues. Sanfilippo A is the most common, and most severe, of the Sanfilippo subtypes with an incidence of up to 1 per 114,000 live births (Sharkia *et al.*, 2014). It has previously

been suggested that the 4 types of MPS III are clinically indistinguishable, but subtype A is more severe, faster progressing and has an earlier onset and worse prognosis, with patients having shorter life spans (Valstar *et al.*, 2008).

In general, babies with MPS IIIA are of normal birth weight and development appears normal (Sharkia *et al.*, 2014). Sanfilippo A is dominated by neuropsychiatric problems and those acquired can be divided into 3 phases. The first phase begins between the ages of 1 and 4 when children show developmental delay and hindered speech (Cleary and Wraith, 1993; Valstar *et al.*, 2008). The second phase begins around 3 years of age and is the phase in which the behavioural problems begin (Cleary and Wraith, 1993). These issues include hyperactivity, restlessness, tantrums, including destructive behaviour, and are generally very difficult to manage (Sharkia *et al.*, 2014). Sleep disturbances are frequently reported in this phase and can include constant waking, difficulty settling, early rising and occasionally complete reversal of circadian rhythms (Valstar *et al.*, 2008). At approximately 10 years of age behavioural issues begin to subside as gradual loss of motor and mental function occurs. This is the final phase of MPS IIIA. Initially, falls are frequent but eventually the patient is wheelchair bound as neurodegeneration continues. Swallowing difficulties can begin to arise and feeding tubes become necessary. Seizures are also common during this phase of the disease (Cleary and Wraith, 1993). Early death occurs in the mid to late teenage years and is often caused by pneumonia (Valstar *et al.*, 2010).

Somatic features are clearly overshadowed by the neurological problems in MPS IIIA, but although mild they are still present. Patients tend to have mild coarse facial features including prominent eyebrows and dry, coarse facial hair which grows abnormally. Lips also appear thick and turned out. Affected individuals appear to have a longer skull and a shorter forehead and macrocephaly is common. The facial dysmorphisms are more obvious in older patients and so it is thought they develop in the latter stages of the disease (Valstar *et al.*, 2008). Hepatomegaly, inguinal and umbilical hernias, scoliosis and kyphosis have also been reported (Valstar *et al.*, 2010; Sharkia *et al.*, 2014).

MPS IIIA results in the accumulation of partially degraded HS, which is often highly sulphated. The aberrant sulphation patterns may lead to alterations in ligand binding and downstream signalling if the abnormal HS is released from the lysosomes (Wilkinson *et al.*, 2012). The build-up of the HS in the lysosome inhibits the function of other lysosomal hydrolases leading to the accumulation of sialylated glycosphingolipids known as gangliosides (Walkley, 2004). Simple gangliosides GM2 and GM3 accumulate within the neurons where they initiate an immune response that contributes to the disrupted neurological phenotype (Fedele, 2015). It is important to note that ganglioside accumulation is seen in a variety of lysosomal storage disorders and not just MPS or MPS IIIA indicating their accumulation is not GAG specific but related to the dysfunction of the lysosome (McGlynn, Dobrenis and Walkley, 2004; Viana *et al.*, 2020). As well as ganglioside accumulation, it has been found that apoptosis and oxidative stress are increased over the progression of the disease. Furthermore, microglial activation is increased indicating neuroinflammation is involved in disease progression, though it is unclear whether neuroinflammation occurs as a result of neurodegeneration, or *vice versa* (Arfi *et al.*, 2011; Wilkinson *et al.*, 2012).

1.5.2.1 Treatment of Mucopolysaccharidosis Type IIIA

Due to the neurological pathology associated with MPS IIIA, any therapies need to have the ability to cross the blood-brain-barrier (BBB). This can be extremely challenging and many therapies that are used for other subtypes of MPS cannot be used for MPS III. Enzyme replacement therapy (ERT) is commonly used for other MPS subtypes but is unable to cross the BBB so the neurological aspects of MPS IIIA are not reversed (Concolino, Deodato and Parini, 2018). The SGSH enzyme can be administered directly into the central nervous system (CNS) but a trial in dogs revealed high enzyme doses were required and the procedure carries high risks (Gaffke *et al.*, 2018). It is possible that the SGSH enzyme can be altered to include an IgG domain which will allow uptake through the BBB by specific receptor binding, but this requires further investigation (Boado *et al.*, 2014). Substrate reduction therapy (SRT) is based on decreasing the biosynthesis of GAGs to reduce the amount of substrate of the deficient enzyme and reduce the accumulation of partially degraded HS. Unfortunately, trials in

humans have revealed no permanent reduction in MPS IIIA symptoms (Fecarotta, Gasperini and Parenti, 2018; Gaffke *et al.*, 2018; Seker Yilmaz *et al.*, 2021).

In vivo gene therapy has shown very positive results in MPS IIIA animal models. The *SGSH* gene is delivered directly to patients so the correct enzyme can be produced, and degradation restored (Beck, 2018). Mouse models have shown that using a recombinant adeno-associated virus (AAV), more specifically AAV serotype 5, to deliver sulphatase modifying factor 1 (SUMF1) directly into the CNS has beneficial effects on the neurological symptoms associated with MPS IIIA (Fraldi *et al.*, 2007). As with other emerging therapies, further work is needed to understand the efficacy and safety of in vivo gene therapy by carrying out more advanced studies on larger groups of patients (Gaffke *et al.*, 2018).

Another potential cell therapy is haematopoietic stem cell (HSC) transplant which involves the transfer of donor bone marrow or umbilical cord cells into the patient. The donor cells begin to differentiate and cross correct cells throughout the body. One major benefit of this therapy is that the cells can cross the BBB and differentiate to correct the CNS and reverse neurological pathophysiology (Langford-Smith *et al.*, 2012). Although not an MPS subtype, metachromatic leukodystrophy (MLD) is a disease caused by the deficiency of Arylsulphatase A, an enzyme involved in the catabolism of sphingolipids. The disease is characterized by significant loss of motor and cognitive functions (Gomez-Ospina, 1993). Recently, gene therapy has been trialled in animal models and resulted in complete correction of the disease pathology (Audouard *et al.*, 2021). The therapy, known as Libmeldy®, involves collection of HSCs from patients which then have a functional gene inserted in place of the mutated one before reinfusing the corrected cells back into the patient (Federico and de Visser, 2021). Libmeldy® has now been approved for use in children and provides exciting potential for other LSDs (National Institute For Health and Care Excellence, 2022). In December 2022, Orchard Therapeutics reported that an HSC gene therapy had been trialled in patients, with promising results. SGSH levels appeared increased in bone marrow derived cells, plasma and cerebrospinal fluid and abnormal HS levels were reduced. Four out

of five patients on the trial showed gain of cognitive skills while three patients remained within the normal development quotient up to 18 months post treatment. Although this is a promising new tool, the patients need further monitoring to assess overall outcomes of the trial (Orchard Therapeutics, 2022).

hiPSCs also present as a promising tool. Somatic cells from MPS IIIA patients can be reprogrammed to hiPSCs, genetically corrected and transplanted back into the patient. Mice with MPS IIIB have undergone this type of treatment where the genetic defect was corrected using a lentiviral vector before the repaired cells were transplanted into the mice intracranially. The study showed promising results with a reduction in the lysosomal pathology, indicating it could be used as a potential treatment in humans (Clarke *et al.*, 2018). The use of hiPSCs as investigative models, as well as their drawbacks for therapy are discussed in greater detail in Section 1.7.3.

Many current gene therapies include the use of lentiviruses and AAVs, both of which have been using in humans with MPS IIIA. Both of these therapies are reported to increase enzyme production in the brain and relieve the clinical features associated with MPS IIIA (Rossi and Brunetti-Pierri, 2023). CRISPR/Cas9 provides an exciting potential gene editing system that can be applied to various diseases. CRISPR/Cas9 gene editing has been used in humans and one study showed promising results in a patient with sickle cell anaemia, but the technique has not yet been applied the GAG-associated diseases (Gostimskaya, 2022). However, the technique is still very expensive and there remains some concerns over its safety and efficacy. One major concern surrounds off-target effects. The CRISPR/Cas9 machinery has to potential to induce double stranded breaks at unwanted locations, which could generate large deletions, rearrangements or even point mutations. These could then lead to a variety of unwanted genetic diseases that the patient did not have prior to the CRISPR/Cas9 therapy (Christensen and Choy, 2017; Benetó *et al.*, 2020; Poletto, Baldo and Gomez-Ospina, 2020; Seker Yilmaz *et al.*, 2021).

1.5.2.2 Human Induced Pluripotent Stem Cell Models of Mucopolysaccharidosis Type IIIA

When stem cells have been used to study MPS disorders, investigators have frequently used hiPSCs reprogramed from patient derived cells rather than the application of gene editing (e.g. CRISPR/Cas9) to create knockouts of lysosomal enzymes. MPS IIIA is the most prevalent subtype of Sanfilippo syndrome, so the abundance of available patient material is much higher compared with other HS associated diseases (Li *et al.*, 2018). The first MPS IIIA patient-derived hiPSC line was reported in October 2018 and was derived from a three-year-old female patient with a compound heterozygous mutation in exon 6 and exon 8 (**Figure 7**). The hiPSC lines were generated from dermal fibroblasts and expressed pluripotency markers (SOX2, OCT4 and NANOG) and were able to differentiate to generate cells representing all three germ layers. The differentiation was carried out by embryoid body formation and one week of spontaneous differentiation in Dulbecco's Modified Eagle Medium (DMEM). The appearance of differentiated cells from all three germ layers was confirmed by the presence of Nestin and TUJ1 (neural), α-actinin and GATA4 (mesoderm), and AFP and SOX17 (endoderm) (Vallejo *et al.*, 2018).

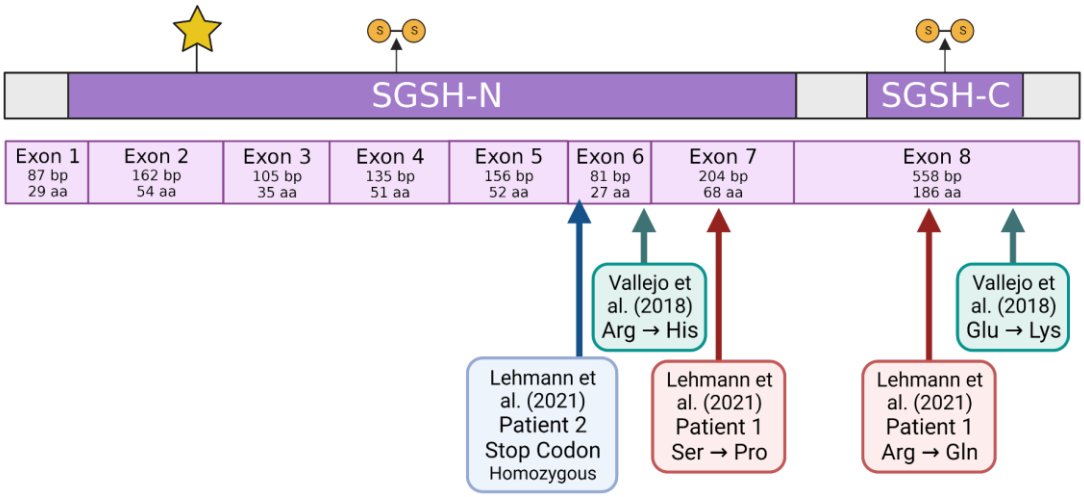


Figure 7 – Mutations of MPS IIIA Patient Derived hiPSCs. The *SGSH* gene is shown with the active site (yellow star), disulphide bonds (yellow circles) which are important for the stability and conformation, exon number, base pair (bp) number and amino acid (aa) number. Patients from two studies are shown, two of which have compound heterozygous mutations that result in amino acid substitution and one patient who has a homozygous mutations resulting in a stop codon.

This finding is in stark contrast to a separate study conducted by Lehman *et al*. (2021), in which two other hiPSC lines were derived from different patients. Patient 1 had an intermediate phenotype caused by a compound heterozygous mutation in exon 7 and exon 8 while patient 2 displayed a much more severe phenotype caused by a homozygous mutation in exon 5 which resulted in a stop codon (**Figure 7**). The SGSH activity of the cells was assessed and the intermediate phenotype hiPSCs activity was 16.4 pmol/g/mg protein and the severe phenotype hiPSCs activity was 0.85 pmol/g/mg protein compared to a control activity of 116.1 pmol/g/mg protein. This indicates that the levels of abnormal HS would be increased in patient 1 compared to control lines, and then further increased in patient 2. Both of these hiPSC lines (from patient 1 and 2) expressed multiple pluripotency markers (NANOG, OCT4, SOX2), yet the hiPSC lines derived from patient 2 were unable to differentiate despite pluripotency levels appearing similar to controls. In contrast, hiPSC lines derived from patient 1 formed neural precursor cells (NPCs) similar to control cell lines. Once formed, the NPCs were cultured in the presence of FGF2 and maintained. It was found that the NPCs derived from patient 1 hiPSCs grew more slowly compared to control hiPSCs, but upon the addition of excess FGF2, the proliferation could be rescued. Further investigations concluded that MPS IIIA hiPSCs were able to bind FGF2 similarly to control hiPSCs, but the response, and downstream signalling was significantly reduced.

After exploring the ability of the hiPSCs to form NPCs, FGF2 was removed to allow the NPCs to differentiate into neurons. It was found that the MPS IIIA hiPSC line had severely inhibited neurogenesis, showed by a significant reduction in neural markers, including *NESTIN* and *βIII-TUBULIN*, when compared with wild type control cell lines. The group note that the creation of isogenic control lines would allow further understanding of the *SGSH* mutations. It is important to note that the neural markers were upregulated in both cell lines compared with the NPCs, yet the markers expressed in MPS IIIA hiPSC derived neurons were reduced compared to controls. Additionally, partially degraded GAG from MPS IIIA hiPSCs was added into control neuron cultures to determine whether the inhibited neurogenesis was a result of

abnormal GAG. The neurogenesis of hiPSCs under these conditions was decreased when compared with untreated lines, indicating that the build-up of abnormal HS accounts for the impaired differentiation (Lehmann *et al.*, 2021).

The findings of the Lehmann *et al.* (2021) study indicates that the impact of *SGSH* mutations is apparent later in *in vitro* differentiation, as NPCs begin to mature to neurons. This could also explain why the impact of the mutation was not picked up in the Vallejo *et al.* (2018) study, as this was a much shorter and simpler differentiation protocol. Furthermore, different mutations may have divergent effects on the differentiation capability of the hiPSCs. This could be due to a number of reasons including the location of the mutation, the number of base pairs affected and how the different amino acid substitutions affect the protein structure and folding. The effect on the protein and gene will impact the amount of abnormal GAG that builds up which can interfere with key signalling pathways. It has previously been reported that multipotent adult progenitor cells (MAPC) from patients with MPS I have disrupted interaction between FGF2 and fibroblast growth factor receptor (FGFR) 1 and also impaired bone morphogenetic protein (BMP) 4 signalling which can be rescued with addition of the deficient enzyme (IDUA) to remove abnormal breakdown products (Pan *et al.*, 2005; Khan *et al.*, 2008). MPS IIIA patients show a range of phenotypes which range from severe to intermediate, so patient derived hiPSCs will also show variation.

As MPS IIIA is characterized by the build-up of abnormal GAG over time, the effect of *SGSH* mutations will likely not be seen in the short term. This is likely the reason Lehmann *et al.* (2021) only saw differences in longer differentiation protocols. For this reason, a much longer differentiation protocol would need to be carried out to determine whether the mutation has any effect on mesoderm and endoderm differentiation as this has not been investigated to date. As the critical disease phenotype is neurodegeneration, it is not surprising that the neural differentiation pathway has received the greatest focus in MPS modelling hiPSC studies. However, investigations of other lineages could provide additional information about the impact of accumulating abnormal HS on other signalling pathways and cell behaviours.

1.6 Methods of Glycosaminoglycan Analysis

GAGs are highly heterogenous and often their composition and sulphation pattern confer their function. Furthermore, GAGs are synthesized by a non-template driven process, so they are much harder to investigate when compared with DNA, RNA and proteins. Methods of analysing GAGs tend to provide either structural information (compositional disaccharide analysis and metabolic radiolabelling) or spatial information (antibodies).

1.6.1 Compositional Disaccharide Analysis

A common method for the analysis of GAGs relies on their cleavage from core proteoglycans, purification and chain depolymerization by enzymatic degradation into oligosaccharides. The enzymes, known collectively as GAG lyases, digest the GAG chains into smaller fragments. These are then separated and detected by high performance liquid chromatography (HPLC), alone or in combination with mass spectrometry (MS) or nuclear magnetic resonance (NMR) (Rice *et al.*, 1985; Deakin and Lyon, 2008; Limtiaco, Jones and Larive, 2009; Volpi and Linhardt, 2010).

One advantage of using GAG lyases is the selectivity they provide. For example, a purified sample will contain a mixture of GAGs, but only those that have been digested will be detectable due to the formation of the unsaturated bond. HS is digested by heparinase I, II and III which cleave the chain between the glucosamine and uronic acid resulting in a mixture of different HS disaccharides. One major drawback of using heparinases for chain digestion is that the stereochemistry of the uronic acid is lost and the difference between IdoA and GlcA cannot be determined. The HS depolymerization results in the formation of a double bond between C4 and C5 of the uronic acid (Archer-Hartmann *et al.*, 2023). Strong-anion exchange (SAX)-HPLC can be used to separate the disaccharides and exploits the formation of the double bond which absorbs UV at 232 nm (Rice *et al.*, 1985; Yang *et al.*, 2011). UV based detection methods are not sensitive for small amounts of GAG and the UV detection method results in the detection of non-GAG species (Deakin and Lyon, 2008).

This thesis uses reverse phase (RP)-HPLC to separate HS disaccharides fluorescently labelled with 2-aminoacridone (AMAC) (**Figure 8**). RP-HPLC provides improved resolution of the disaccharides and the hydrophobic nature of the AMAC increases sensitivity. When compared with SAX-HPLC, RP-HPLC is a much more common technique and is readily available in most laboratories, so is often the method of choice for GAG analysis. In comparison to other fluorescent labelling techniques, free AMAC is retained by the column and so excess does not need to be purified out from the sample prior to HPLC separation. The excess AMAC is removed from the column during washing. AMAC labelled samples are run alongside known standards so the disaccharides in the sample can be determined based on elution time. An example of HS disaccharide standard RP-HPLC separation can be seen in **Appendix 3**. This provides information on the contribution of each disaccharide to the overall GAG population detected. It is important to note that labelling efficiency may vary, so the amount detected does not always correlate with the amount that is actually present in the sample (Deakin and Lyon, 2008). However, correction factors can be determined (the equation for this can be seen in section 4.2.5) and applied to the peaks to determine the amount of each disaccharide. During the time of this project, 3-O-sulphated disaccharide standards were not available and so the 3-O-sulphated disaccharides were not detectable. More recently, a set of 3-O-sulphated standards have been generated by synthesising tetrasaccharides which were subsequently sulphated at the 3-O position and digested into disaccharides. Eight disaccharides were generated with varying degrees of sulphation. The disaccharides were able to be fluorescently labelled and purified on an HPLC and their structures confirmed by MS (Karlsson *et al.*, 2021). This is a great advancement and will provide the ability to detect more sulphation patterns within the HS chain.

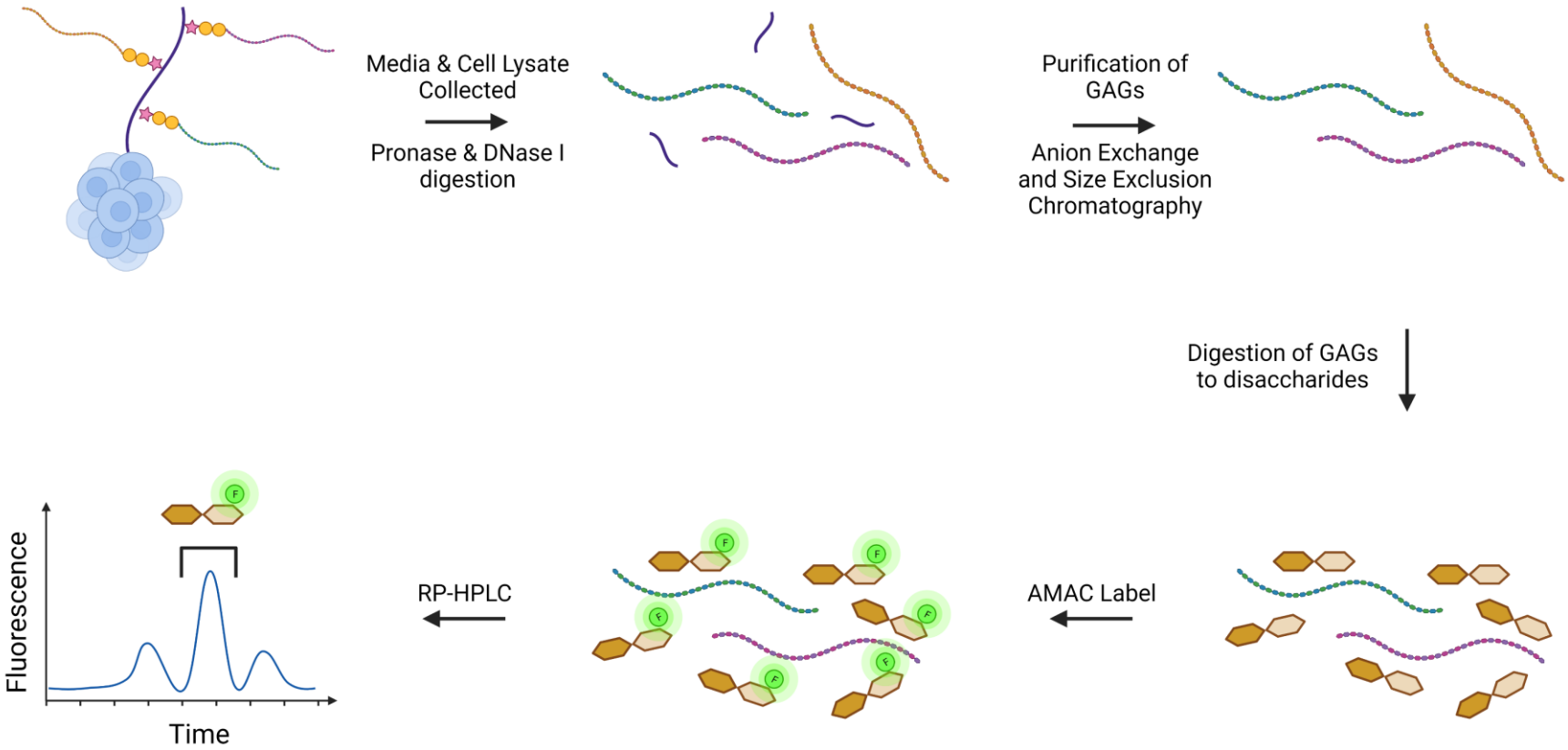


Figure 8 – Compositional GAG Disaccharide Analysis by AMAC-labelling RP-HPLC. Cells are cultured under standard pluripotency conditions (E8 and VTN-N). Conditioned medium (containing excreted PGs) is collected and cells are lysed (to collect membrane bound PGs). Pronase is used to digest GAGs away from PG core and DNase I digests DNA from cell lysates. Full length GAG chains are purified using anion exchange chromatography and size exclusion chromatography. Specific GAGs are depolymerized using their relevant GAG lyase and disaccharides are fluorescently labelled using AMAC. Labelled disaccharides are separated and detected using RP-HPLC using known standards for reference.

1.6.2 Glycosaminoglycans Antibodies

Currently, two classes of antibody are available to detect HS – hybridoma derived and phase display derived (David *et al.*, 1992; Van Kuppevelt *et al.*, 1998). The hybridoma derived antibodies are commercially available: 10E4 IgM and 3G10 IgG (**Figure 9**). 10E4 recognises N-sulphated HS and due to the nature of the sulphated domains, and IgM type antibody, it is thought that 10E4 has multiple binding sites throughout the chain. If the HS is treated with heparinase III, 10E4 can no longer bind and the HS cannot be visualized. In contrast, 3G10 recognizes a neo-epitope on HS chains which is only exposed after treatment with heparinase III and so will not bind to untreated HS. Only one 3G10 antibody can bind per epitope, so the staining can be quantified to indicate the number of HS chains present in a sample. 3G10 and 10E4 are highly specific to HS and do not recognise any other GAGs (David *et al.*, 1992).

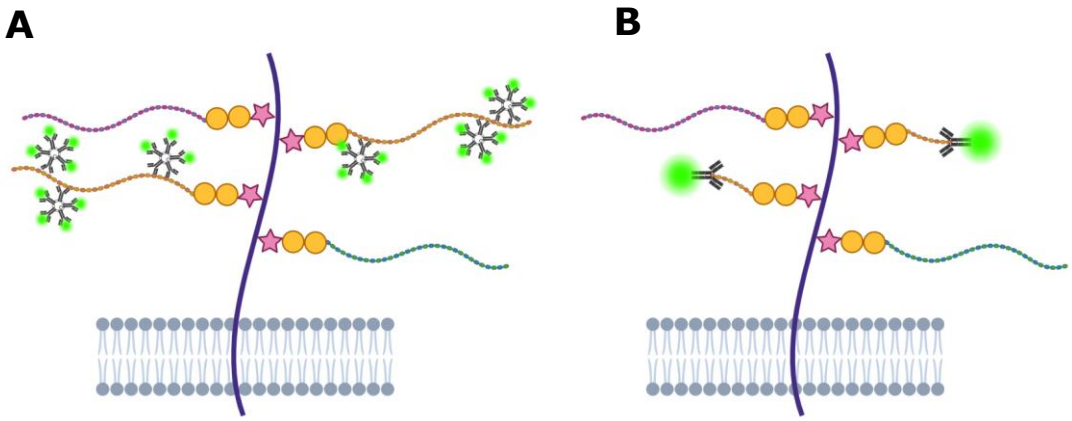


Figure 9 - Hybridoma Derived GAG Antibodies. A) 10E4 is an IgM antibody that recognises N-sulphated HS and binds several regions throughout the chain. B) 3G10 recognises a neo-epitope on HS chains that have been digested with heparinase III. The IgG antibody only binds to one epitope so the fluorescence can be quantified and used to indicate the number of HS chains.

Phage display derived antibodies have been generated by the Van Kuppevelt lab. The antibodies were raised against HS derived from animal tissues and so use natural HS. The antibodies are single-chain variable fragments (ScFv) and so rather than being

traditional antibodies, they are formed from variable regions of heavy and light chains connected by linker sequences (Smits *et al.*, 2006). The ScFv antibodies recognise different sequences of the HS, but the specific epitopes cannot be determined for all of the ScFvs and instead they recognise a range of disaccharides or sulphation modifications. **Table 3** shows the preferred binding epitopes for each of the ScFvs. Throughout this project 10E4 has been used to detect HS and AO4B08 was used to detect 2-O-sulphated HS as antibodies for endogenous levels of HS2ST1 are not available. AO4B08 has been used previously for this purpose and its application is well described (Johnson *et al.*, 2007; Kurup *et al.*, 2007; Qiu *et al.*, 2018)

Table 3 - Heparan Sulphate Phage Display Derived ScFv Antibodies

Antibody	Preferred Epitope	Reference
AO4B08	GlcNS±6S-IdoA2S-GlcNS±6S	(Kurup *et al.*, 2007)
RB4Ea12	GlcNS6S-IdoA-GlcNS6S	(Jenniskens *et al.*, 2002)
HS4E4	NS, NAc, IdoA	(Kurup *et al.*, 2007)
NS4F5	$(NS,6S\text{-}IdoA,2S)_3$	(Smits *et al.*, 2010)
EV3C3	NS, 2S, IdoA	(Thompson *et al.*, 2009)
HS4C3	NS, 2S, 3S, 6S	(Van Kuppevelt *et al.*, 1998)
LKIV69	NS and 2S	(Wijnhoven *et al.*, 2006)

1.6.3 Time of Flight Secondary Ion Mass Spectrometry

Time of flight secondary ion mass spectrometry (ToF-SIMS) is a surface analysis tool which, unlike the methods described above, provides both spatial and compositional information. SIMS involves firing an ion beam at the surface of a sample to generate secondary ions which are then picked up by a detector (Vickerman, 2013).

To date, ToF-SIMS has been able to distinguish between pharmaceutical heparin from different sources indicating it is highly sensitive. Furthermore, ToF-SIMS is able to detect small amounts of GAGs and samples do not require digestion or labelling. Although ToF-SIMS presents as a potential tool for high-throughput analysis of GAGs, the mass resolving power is low, so ions of similar mass value cannot be distinguished (Hook *et al.*, 2021). ToF-SIMS can be combined with an Orbitrap mass analyser which provides a much better mass resolution, yet the spatial resolution is poor when compared with ToF-SIMS (Passarelli *et al.*, 2017). The combined use of these two techniques (known as 3D-Orbi-SIMS) provides a novel tool for the spatial and compositional analysis of GAGs. hiPSCs lacking HS were initially compared with WT hiPSCs using ToF-SIMS and differences were seen, indicating this method can be taken further (Jamie Thompson, personal communication). Since this discovery, the use of ToF-SIMS and 3D-Orbi-SIMS have been used to identify different GAGs in a range of tissues and cell lines, but the technique is still under development (Lorna Milne, personal communication). One major drawback of using 3D imaging mass spectrometry is the matrix effect. The matrix effect refers to differences in ionization between repeated samples due to the surrounding matrix that can either inhibit or enhance the formation of certain ions. For example, increased intensity of ions does not always mean increased concentration (Shard *et al.*, 2015). In addition, biological validation is needed with every run, so standards or controls need to be included.

1.7 Human Induced Pluripotent Stem Cells

Stem cells are present throughout the adult body for cell and tissue regeneration. These stem cells are multipotent meaning they can only differentiate into a limited number of cell types (Ratajczak *et al.*, 2012). Mesenchymal stem cells (MSCs) are an example and can give rise to bone, cartilage and muscle cells (Caplan, 1991). Pluripotent stem cells (PSCs) are much less restricted and can differentiate to any somatic cell type. For this reason, PSCs provide an invaluable tool for in vitro studies. Embryonic stem cells (ESCs) are an example of PSCs and were first isolated from mouse blastocysts (Evans and Kaufman, 1981). Much of the early work on mESCs improved understanding of pluripotency, optimal growth

conditions, growth factors and gene expression (Ginis *et al.*, 2004). More recently, ESCs from other species, including humans (hESCs), have been isolated. The hESCs were isolated in similar manner, from a blastocyst, and in-depth studies into the survival and pluripotency were carried out. hESCs provide a better tool for studying human disease when compared with mESCs (Thomson *et al.*, 1998). However, the use of hESCs is surrounded by ethical issues due to the requirement of using human embryos as a source.

In 2006, Takahashi and Yamanaka described the induction of PSCs cells from human fibroblasts by the addition of four transcription factors (Oct3/4, Sox2, c-Myc and KLF4). The resulting hiPSCs show similarities to ESCs regarding their morphology and proliferation (Takahashi and Yamanaka, 2006). hiPSCs still provide most of the benefits of hESCs regarding pluripotency, but the ethical issue surrounding the use human embryos is removed. As hiPSCs can be generated from somatic cells, the collection of material is much easier and less invasive (Yamanaka, 2007). hiPSCs can be generated from urine, which contains live cells shed from the urinary system. This means that hiPSCs generation can be completely free of invasive procedures (Zhou *et al.*, 2011).

1.7.1 Human Induced Pluripotent Stem Cells for Disease Modelling

Mouse models are commonly used to study human diseases but as a model there are many disadvantages. The most obvious difference is that mice and humans are different species and while some of the genome is shared, there is still critical variation. Mouse models have low predictive ability regarding human toxicity to drugs and many drugs that have not had an undesirable effect in mice have exhibited toxicity in humans (Uhl and Warner, 2015). In addition, mice are often used to study late onset or progressive diseases which are induced in the mice at a young age. This is not representative of diseases that progressively get worse or are onset later in life. Furthermore, the husbandry conditions can have a large impact on experiments and conditions can vary between laboratories. Even factors such as bedding, and light can affect the mice. Mice in different laboratories will also have a different microbiome and there is mounting evidence to suggest the

microbiome has a huge effect on results; the immune system and neural function is vastly influenced by the microbiome (Justice and Dhillon, 2016).

hiPSCs provide a great potential for the future of *in vitro* studies as an unlimited number of cells can be produced from patient derived iPSCs. Several studies have been conducted to model diseases, identify therapeutic targets and carry out drug screens (Wu and Hochedlinger, 2011). Spinal muscular atrophy (SMA) has been modelled using hiPSCs derived from patients afflicted with SMA. The hiPSCs were differentiated down neural lineages and a progressive loss of motor neurons was seen, a characteristic also seen during disease progression in humans. The loss of neurons in the iPSCs was treated using a compound that increased levels of the protein affected by the genetic mutation causing SMA. The phenotype of the hiPSCs appeared reversed and the loss of motor neurons was reduced (Ebert *et al.*, 2009). This is not the only study to have been conducted, Wu and Hochedlinger (2011) reviewed several studies that build disease models for drug screening and all of the models show a phenotype similar to that of the patient they are derived from (Wu and Hochedlinger, 2011).

As hiPSCs appear to be a great tool for disease modelling and therapeutic identification, their use for cell therapy has also been investigated. An iPSC line derived from a mouse with MPS IIIB (*Naglu*$^{-/-}$) was altered to overexpress *Naglu* using a lentiviral vector. The iPSCs were then differentiated into neural stem cells (NSCs) and grafted into newborn *Naglu*$^{-/-}$ mice. After 9 months, mice showed increased Naglu activity in the brain and a reduction in neuropathological markers, indicating that the engraftment of NSCs overexpressing deficient enzymes could provide a potential therapy option (Clarke *et al.*, 2018). Although this study was only proof of concept, it still holds great promise that genetically corrected iPSCs can be used for cell therapies in the future.

Differentiation of hiPSCs can be, structurally, quite simple (2D and one cell type) so may not represent multicellular diseases well. Organoids are 3D cell structures that can be generated from hiPSCs. Organoids can be cultured indefinitely and have multiple

cell types within their self-organizing structure. Therefore, they provide a promising tool for investigating more complex diseases (Fatehullah, Tan and Barker, 2016; Mollaki, 2021).

1.7.2 Heparan Sulphate and Human Induced Pluripotent Stem Cells

Several factors and signalling pathways help to maintain pluripotency and drive differentiation of hiPSCs *in vitro* including FGF2/PI3K/AKT, MEK/ERK, TGFβ/Activin A/Nodal and Wnt/β-catenin (**Figure 10 and 11**). Many of the factors and signalling pathways require careful regulation as pluripotency is a precise balance between opposing signals. Furthermore, the same signals may be present in both pluripotency and differentiation but are at different levels or gradients. It is well reported that the signalling factors involved in differentiation and pluripotency interact with HS. It is thought that HS sequesters growth factors and acts as a reservoir to create concentration gradients. Additionally, some factors, such as FGFs, use HS as a coreceptor and the binding of HS, FGF and the FGFR are thought to be needed to elicit some downstream responses (Hynes, 2009).

FGFs are a family of 22 polypeptides known to be involved in regulating essential biological processes (Mossahebi-Mohammadi *et al.*, 2020). FGF2 binding to its relevant receptor activates the PI3K/AKT and MEK/ERK signalling pathways. When FGF2 concentrations are low, pluripotency is maintained by MEK/ERK signalling. In higher concentrations of FGF2, PI3K activates AKT which, in turn, inhibits MEK/ERK activity. However, in the absence of FGF2 and PI3K signalling, ERK is increased above a threshold, and differentiation occurs. When FGF2 levels are low, the threshold of ERK to initiate differentiation is not reached (Apáti, Berecz and Sarkadi, 2016). ERK also inhibits a glycogen synthase kinase (GSK3β) which decreases c-Myc phosphorylation and improves the stability of c-Myc. There are several pools of GSK3β which are involved in different signalling pathways. ERK inhibits GSK3β involved in Wnt signalling so when ERK is not inhibited by AKT, GSK3β is inhibited by ERK which allows for the Wnt/β-catenin pathway to become activated and initiate differentiation (Singh *et al.*, 2012).

The TGFβ family is also important for pluripotency. The TGFβ ligands, Activin A and Nodal, transcriptionally activate NANOG by SMAD 2/3 signalling (Vallier, Reynolds and Pedersen, 2004; Vallier *et al.*, 2009; Apáti, Berecz and Sarkadi, 2016). FGF2 signalling and TGFβ signalling are not enough to maintain pluripotency alone, and knockouts of one or the other result in differentiation (Vallier, Alexander and Pedersen, 2005). This crosstalk mechanism works by the action of PI3K which restricts SMAD 2/3 phosphorylation to help maintain pluripotency. When PI3K is reduced (as a result of reduced FGF2), SMAD 2/3 phosphorylation is increased and the target genes for activation are switched from pluripotency genes to differentiation genes. Additionally, Wnt signalling allows SMAD 2/3 activity to switch to differentiation genes from those involved in pluripotency (Dalton, 2013).

Wnt is a signalling molecule that interacts with a G protein coupled receptor (GPCR) known as frizzled (Fz). Fz then interacts with dishevelled (DVL) which elicits a downstream response that results in β-catenin translocating to the nucleus. The data surrounding Wnt is conflicting, but mounting evidence suggests that in murine systems, Wnt has a downstream pluripotent effect while in humans Wnt signalling appears to induce differentiation (Apáti, Berecz and Sarkadi, 2016). When investigating Wnt signalling and PI3K signalling together, it is clear that separate pools of GSK3β must exist. In PI3K signalling GSK3β is inhibited while in Wnt signalling GSK3β is activated, with both resulting in pluripotency (Singh *et al.*, 2012; Dalton, 2013). Wnt-associated GSK3β inhibits β-catenin translocation to the nucleus. When FGF2 signalling is reduced, ERK is no longer inhibited by AKT meaning ERK can inhibit Wnt associated GSK3β allowing β-catenin activation and canonical Wnt signalling. In this signalling pathway, β-catenin interacts with SMAD 2/3 to activate differentiation related gene expression (Dalton, 2013).

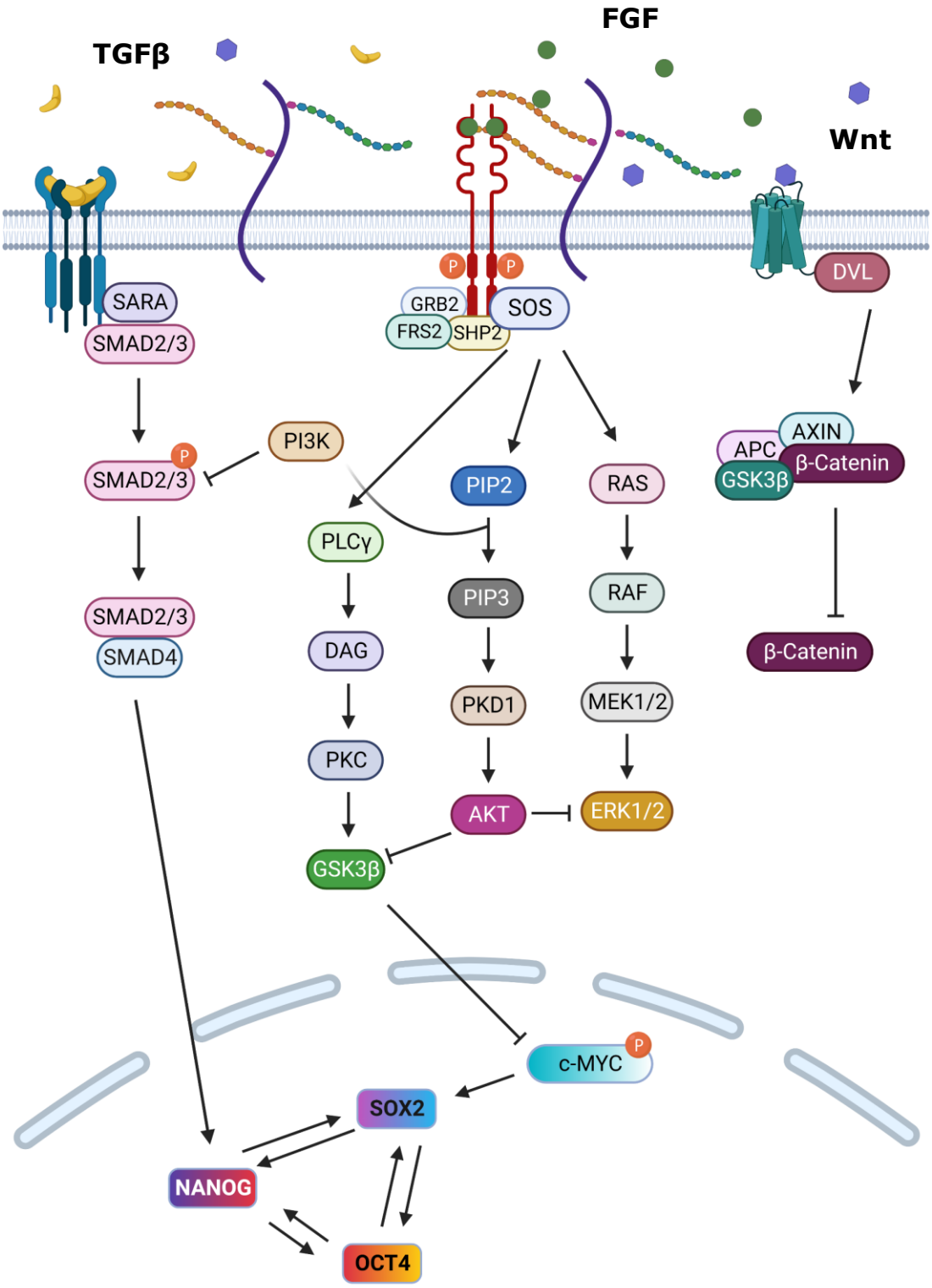


Figure 10 - Pluripotency Signalling in hiPSCs. Pluripotent signalling in hiPSCs is controlled by TGFβ and FGF2 signalling pathways. The inhibition of the Wnt signalling pathway also help to maintain pluripotency by inhibiting differentiation. These signalling factors are known to interact with GAGs and PGs. TGFβ induces pluripotency by acting though SMAD 2/3 signalling. FGF2 acts through PI3K/AKT which in turn inhibits MEK/ERK signalling.

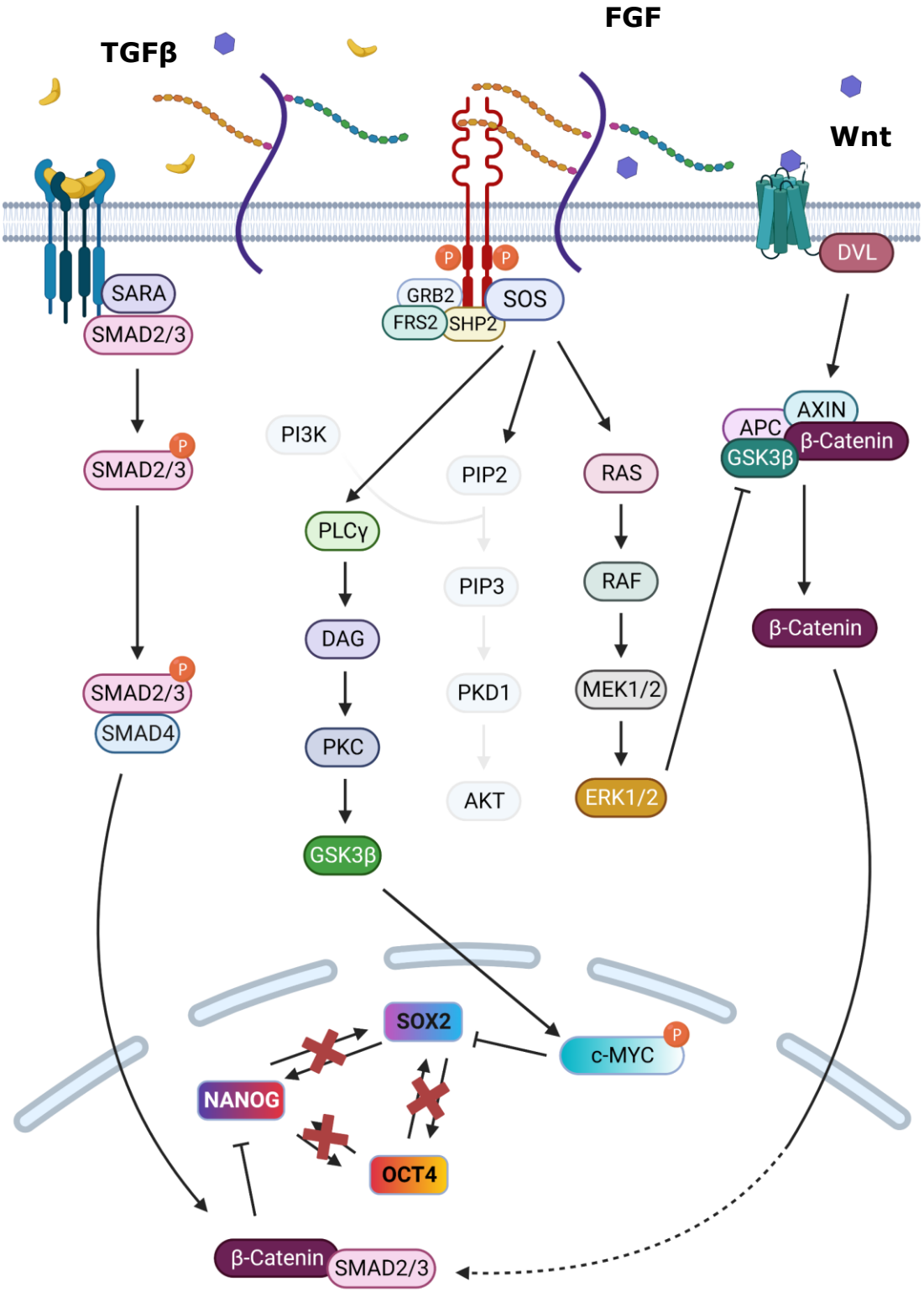


Figure 11 - Differentiation Signalling in hiPSCs. Pluripotency and differentiation signalling are very similar with TGFβ, FGF2 and Wnt signalling pathways all involved. These signalling factors are known to interact with GAGs and PGs. TGFβ induces differentiation by acting though SMAD 2/3 signalling. PI3K/AKT is inhibited and MEK/ERK signalling is activated. Downstream cascade of Wnt signalling is also activated and inhibits genes associated with pluripotency.

1.7.3 Challenges with Human Induced Pluripotent Stem Cells

Although hiPSCs are invaluable resources for research and have the potential for clinical applications, there are still many social, ethical and legal issues that need to be addressed. Their genetic stability is a major concern as genetic disruption can occur during reprogramming, or after time in *in vitro* culture (Laurent *et al.*, 2011). These alterations may affect the proliferation and differentiation capacity of the cells and can increase their tumorigenicity – making them unsafe for transfer back into patients (Zheng, 2016). In addition, hiPSCs that will be used for clinical applications need to be vigorously tested to ensure they are clinically safe and laboratories producing them will need to follow Good Manufacturing Practices (GMP). The conditions of culture will need to be assessed and all laboratories generating hiPSCs for clinical applications will need to rigorously follow them (Moradi *et al.*, 2019).

hiPSCs require a matrix to attach to during *in vitro* culture. Popular matrices are Matrigel® and laminin which are both obtained from mouse Engelbreth-Holm-Swarm (EHS) cells. This method of preparation not only requires the use of animals, but also presents batch-to-batch variation and the exact composition is unknown. In 2011, Chen *et al.* described a novel method for hiPSC culture which is fully defined and uses Essential 8 (E8) medium and recombinant vitronectin-N (VTN-N) (Chen *et al.*, 2011). Other defined matrices, such as recombinant laminin, can also be used, both for pluripotency and differentiation conditions. The culture conditions are particularly important in the present study as many animal-derived products, such as Matrigel®, contain exogenous GAGs. These exogenous GAGs could interfere with the analysis of GAGs produced by the hiPSCs which, in this study, have been genetically altered to produce abnormal HS.

1.8 CRISPR/Cas9 Gene Editing

Gene editing is a technique that is becoming increasingly popular due to its potential as a disease modelling, drug screening and gene therapy tool. There are three main techniques that can be used for gene editing: zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and CRISPR/Cas9.

CRISPR/Cas9 is an easier, more efficient, and simpler method of genome editing when compared with ZFNs and TALENS (Gupta and Musunuru, 2014; LaFountaine, Fathe and Smyth, 2015). Since its discovery as a potential method of gene editing in 2012, CRISPR/Cas9 has been used, and developed, extensively (Jinek *et al.*, 2012).

CRISPR is derived from the immune system of prokaryotes and can be categorised into six distinct groups. Type II CRISPR systems are those that have been harnessed for genome engineering and utilize one endonuclease, Cas9, and two guide RNAs: CRISPR RNA (crRNA) and trans-acting CRISPR RNA (tracrRNA) (Kim, Kang and Ju, 2017). The crRNA and tracrRNA hybridize to a guide RNA (gRNA) and then form a complex with the Cas9, known as the ribonucleoprotein (RNP) complex. The tracrRNA provides the binding site for the Cas9 while the crRNA directs the complex to complementary target sites in the genome where Cas9 can cleave the DNA to induce a double stranded break (DSB), providing the site is adjacent to sequences known as protospacer adjacent motifs (PAMs) (Sander and Joung, 2014). The crRNA is the sequence that can be customized for genome editing in the laboratory.

DSBs can be repaired by one of two methods: homology directed repair (HDR) or non-homologous end joining (NHEJ) (**Figure 12**). HDR relies on a donor template with flanking sequences complementary to the host DNA, these are known as arms of homology. The DSB is repaired as per the donor DNA meaning HDR can be used to insert specific mutations, sequences, or repairs. Although HDR has many obvious benefits, it is highly inefficient. NHEJ does not require a template and relies on the hosts own DNA repair mechanism. As DNA repair is often error prone, NHEJ regularly results in the introduction of insertion/deletion mutations (indels). These indels can give rise to early stop codons and frameshift mutations which can cause proteins to be non-functional or partially functional (Cortez and Patrick, 2015; Wyatt and Ramsden, 2015).

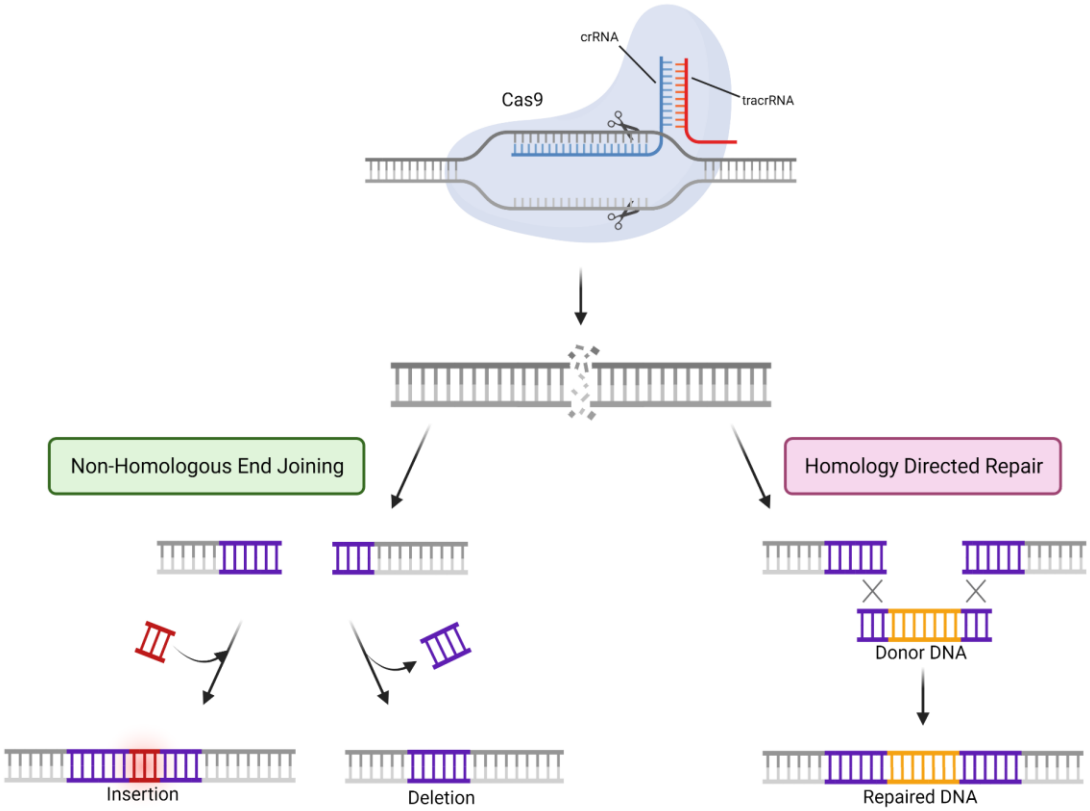


Figure 12 - Mechanisms of DNA Repair After Cas9 Double Stranded Break. After a double stranded break non-homologous end joining (NHEJ) uses the cells intrinsic DNA repair mechanisms which is highly error prone. Homology-directed repair uses donor DNA to induce specific DNA alterations.

1.8.1 Methods of CRISPR/Cas9 Delivery

The CRISPR/Cas9 machinery can be delivered to cells in three ways: DNA, RNA, and protein (**Figure 13**). Each method has its own advantages and disadvantages, but the best method will vary based on the cell type and gene targeted.

DNA delivery of the CRISPR/Cas9 machinery relies on plasmids. Restriction enzymes are used to digest the plasmids before they are ligated with a specific oligonucleotide (Liu *et al.*, 2017). The Cas9, gRNA and HDR templates can be delivered to the cells as individual plasmids or can all be ligated into the same plasmids. Either way, the resulting plasmids are quite large (Wang *et al.*, 2019). Once the plasmids have been generated, they are delivered to the cells where they must cross the cell membrane and the nuclear membrane where they need to be digested and the DNA

transcribed and translated. The time between plasmid delivery and genome editing is often much longer compared to other methods and can result in greater off-target effects due to prolonged exposure to the Cas9 (Liu *et al.*, 2017). The DNA and plasmids are often very stable meaning they're much easier to handle in the laboratory, yet this can also amplify the issues with off-target effects once inside the cells. Plasmid based delivery is relatively low cost and easy to handle, so is a common method of choice (Wang *et al.*, 2019).

RNA delivery involves the introduction of the gRNA and Cas9 mRNA which is then translated by the ribosomes prior to RNP formation (Xu *et al.*, 2019). mRNA delivery significantly reduces off-target effects due to transient expression of the Cas9. However, this can result in reduced genome editing efficiency. Furthermore, the mRNA is highly susceptible to degradation by RNases and so this method requires appropriate preventative measures (Wang *et al.*, 2019). In order to form the RNP complex, the Cas9 protein and gRNA must be present in the cells at the same time. As the Cas9 mRNA first needs to be translated, the delivery of the gRNA 6 hours after the Cas9 mRNA has been found to improve efficiency (Jiang and Doudna, 2017).

CRISPR/Cas9 machinery can also be delivered to the cells in protein form. The formation of the RNP complex inside the cells can be exploited and the process done prior to delivery (**Figure 13C**). This method circumvents the need for transcription and translation of either DNA or RNA and the RNP complex can begin to edit the genome immediately. RNP delivery is the most direct and fast-acting method of genome editing and has also been shown to have a high editing efficiency. Furthermore, the RNP has the most transient expression time which further reduces off-target effects (Wang *et al.*, 2019). As the complex is formed prior to delivery, the Cas9 acts as a protector to the gRNA to prevent degradation (Xu *et al.*, 2019). Additional benefits of this method include decreased immune response and decreased cytotoxicity, which can occur with DNA and RNA delivery (Liu *et al.*, 2017). The Cas9 protein must be extremely pure which is much more difficult to achieve but is available for purchase (Wang *et al.*, 2019). For this reason, RNP

delivery is more expensive than the other two delivery methods, but the disadvantages and risks are significantly reduced.

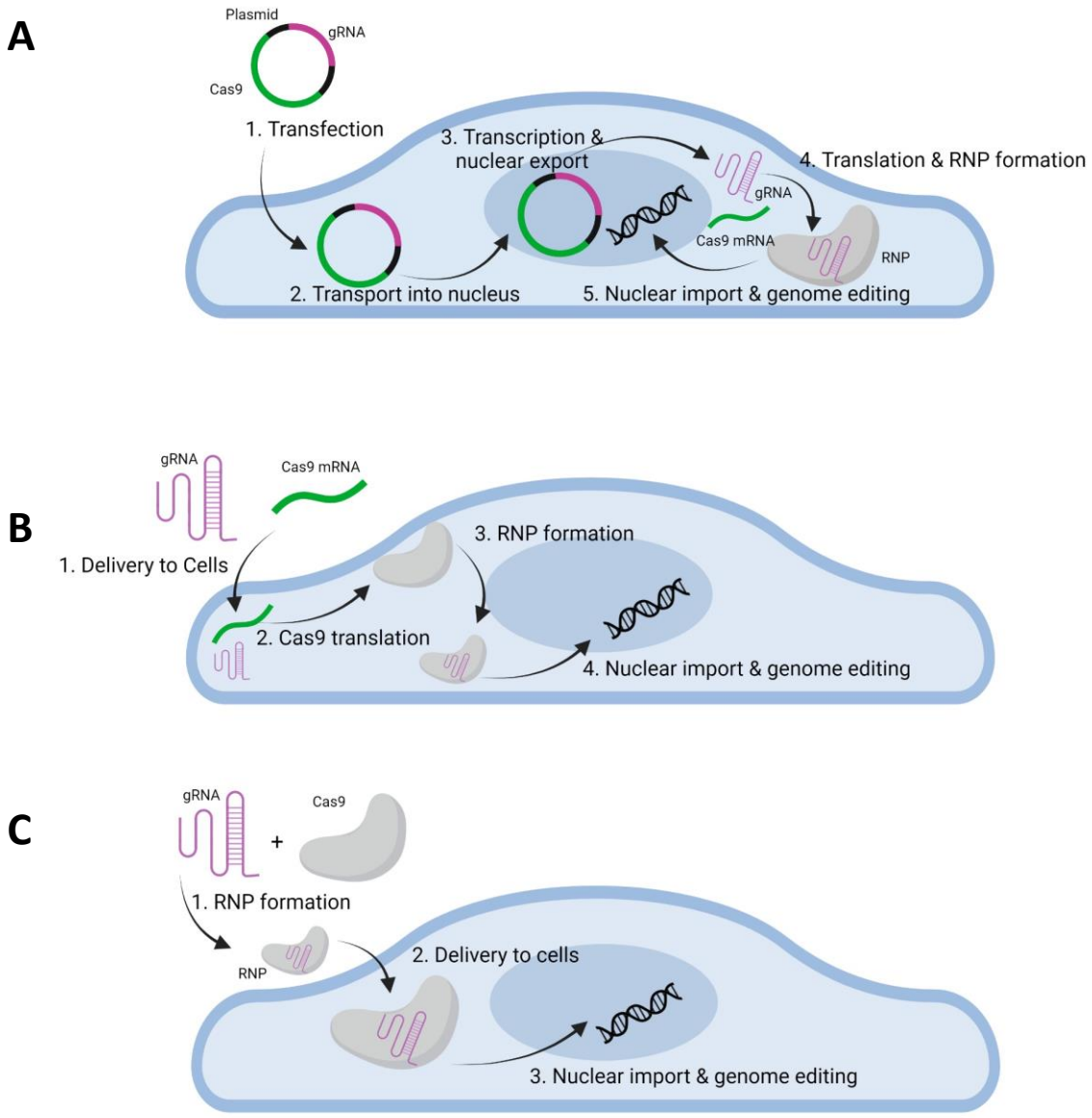


Figure 13 - Methods of CRISPR/Cas9 Machinery Delivery. Cas9 and guide RNAs can be delivered to cell by A) DNA, B) RNA or C) protein delivery. DNA delivery relies on plasmid generation and delivery which is then broken down and transcripted/translated in the cells. RNA delivery requires Cas9 translation in the cells before RNP formation inside the cells. RNP delivery forms the RNP complex outside of the cells which is then delivered and can begin genome editing immediately. (RNP; ribonucleoprotein)

1.8.2 CRISPR/Cas9 and Glycosaminoglycans

Due to the many benefits of CRISPR/Cas9 to study the function of different biological molecules and diseases, it has now become more popular for GAGs. GAGs are incredibly complex and although sulphation patterning is related to function, there are still many unknowns. Additionally, many of the GAG biosynthetic enzymes have several isoforms with similar sequence identity. For this reason, CRISPR can be effective as it is highly precise, but off-target effects need to be closely monitored. One of the first big steps towards a fuller understanding of GAG biosynthesis and structure was the creation of a library of CHO cells with varying CRISPR/Cas9 knockouts or knockins. This library provided insight into the role of each biosynthetic enzyme in the HS and CD/DS pathways as the resulting GAG structures were characterized as well as their binding capacities (Chen *et al.*, 2018).

Another use for CRISPR/Cas9 in the GAG field is for MPS research and treatment. CRISPR/Cas9 provides a route for potential therapy for MPS as other options are limited. As many of the mutations that cause MPS are point mutations resulting in an amino acid change, the DNA sequences can easily be altered by HDR (de Carvalho *et al.*, 2018; Seker Yilmaz *et al.*, 2021). This therapy is still only experimental, and its efficacy and safety need more investigating.

1.9 Project Summary

This PhD project initially set out to understand the effects of abnormal HS degradation on human development. It was hypothesized that hiPSCs with defective HS degradation would have alterations in signalling pathways that involve HS and therefore altered differentiation potential, particularly to the neural lineage. To test this, CRISPR/Cas9 gene editing of the *SGSH* HS degradative gene would be used to generate knockout hiPSCs which could then be directed into differentiation of the three germ layers. This would provide understanding about how the build-up of abnormal HS in the lysosomes of different cells affects cell behaviour.

Despite the COVID-19 pandemic, novel mutations in the HS biosynthetic gene *HS2ST1* were discovered in humans in 2020. In light of this discovery, the focus of the PhD project was changed. It was hypothesized that hiPSCs lacking HS2ST1 function would have increases in other sulphation modifications as well as altered FGF2 signalling. Furthermore, due to loss of an essential HS modification and altered FGF2 signalling, it was hypothesized KO hiPSCs would have altered pluripotency and abnormal differentiation to all lineages. To test this hypothesis, knockout hiPSCs would be generated by RNP delivery of CRISPR/Cas9 machinery. Knockout cells lines would first require confirmation of loss of HS2ST1 function and hiPSCs could then be stimulated to differentiate to each of the three germ layers. An appropriate neural differentiation protocol needs to be optimized and tested as part of this project. The final outcomes and experimental summary can be found in Chapter 6.

Chapter 2

Materials and Methods

2.1 Human Induced Pluripotent Stem Cell Culture

2.1.1 Vitronectin Coating

hiPSCs were routinely cultured on vitronectin recombinant human protein (VTN-N; ThermoFisher, A14700). To coat plasticware, VTN-N stock was diluted in 1X Dulbecco's phosphate buffered saline (DPBS; ThermoFisher 14190094) to achieve a final coating concentration of 0.5 µg/cm^2. In the present study, VTN-N was diluted 1:200 and immediately added to the plasticware at the required volume (**Table 4**) and left at 4°C overnight or room temperature (RT) for at least 1 hour. Prior to the addition of cells or media, excess coating solution was removed. All cultureware used was Nunc coated and T25 flasks were used for routine hiPSC culture.

2.1.2 Human Induced Pluripotent Stem Cell Maintenance

REBLPAT polyclonal hiPSCs (derived from a healthy male, kindly provided by Dr Diogo Mosquiera (Mosqueira *et al.*, 2018)) were routinely cultured in Essential 8 (E8) medium (ThermoFisher, A1517001) which was changed daily. For maintenance, cells were passaged every 72 hours by incubating with TrypLE™ Express Enzyme (ThermoFisher, 12604021) for 3 minutes at 37°C (see **Table 4** for volumes). Excess enzyme was removed, and cells collected in warm E8 medium supplemented with 10 µM Y-27632 (Rho/ROCK pathway inhibitor, ROCKi; Stem Cell Technologies, 72307). Cells were seeded into new plasticware at roughly 20,000 cells/cm^2 in fresh E8 medium supplemented with 10 µM ROCKi. After 24 hours, and every 24 hours thereafter, medium was changed to E8 medium without ROCKi. Cells were cultured at 37°C, in 5% CO_2 in a humid incubator.

Table 4 - Cell Culture Volumes

Vessel	Size (cm^2)	VTN-N Volume	Media Volume (mL)	TrypLE™ Volume (µL)
T25 flask	25	5 mL	5	1000
6 well plate	9.6	2 mL	2	400
12 well plate	3.5	700 µL	1	150
96 well plate	0.32	70 µL	0.1	20

2.2 Immunostaining

One the day of cell collection, medium was removed and cells were washed twice with 1X DPBS prior to fixing with 4% paraformaldehyde (PFA; Polysciences, 18814) for 20 minutes at RT. Cells were washed with 1X DPBS and blocking buffer (1% Bovine Serum Albumin (BSA; Sigma, 232-936-2) in 1X DPBS) was used to block non-specific antibody binding for 1 hour at RT . 0.1% Triton X-100 was added to blocking buffer where nuclear staining was required. Cells were incubated with primary antibody diluted in blocking buffer overnight at 4°C and subsequently washed three times in 1X DPBS. For the ScFv antibody, a secondary antibody (Anti-VSV-G) was diluted in blocking buffer and incubated with samples for 1 hour at RT in the dark. Fluorescent antibodies were diluted in blocking buffer and incubated with cells for 1 hour at RT in the dark before washing three times in 1X DPBS. Antibody details and dilutions can be found in **Table 5**. For nuclear staining, Hoechst 33342 (ThermoFisher, 62249) was diluted to a working concentration of 1 µg/mL in 1X DPBS and incubated with cells for 15 minutes at RT in the dark. Samples were washed twice in 1X DPBS before visualization and imaging on a Leica TCS SPE spinning confocal microscope.

Table 5 - Immunostaining Antibody Details

Antibody	Isotype	Dilution	Supplier Details
10E4	Mouse, IgM	1:200	Amsbio, 370255-1
AO4B08	ScFv	1:20	Gift from Toin H van Kuppevelt (Jenniskens *et al.*, 2000)
NANOG	Mouse, IgG2	1:400	Millipore, MABD24
OCT4	Rabbit, IgG	1:200	Abcam, 19857
NESTIN	Mouse, IgG1	1:500	Abcam, 6320
PAX6	Rabbit, IgG	1:100	Biolegend, 901301
SOX1	Rabbit, IgG	1:100	Abcam, AB109290
β-III-TUBULIN	Rabbit, IgG	1:400	Abcam, 18207
FOXA2	Goat, IgG	1:200	R&D Systems, AF2400
SOX17	Goat, IgG	1:200	R&D Systems, AF1924
Anti-VSV-G	Mouse IgG1	1:1000	Sigma, SAB4200695
AlexaFluor™ 488	Goat anti-mouse IgM	1:1000	Invitrogen, A21042
AlexaFluor™ 488	Goat anti-mouse IgG	1:1000	Invitrogen, A11029
AlexaFluor™ 488	Goat anti-rabbit IgG	1:1000	Invitrogen, A11034
AlexaFluor™ 488	Donkey anti goat IgG	1:1000	Invitrogen, A11055
AlexaFluor™ 594	Goat anti-mouse IgG	1:1000	Invitrogen, A11032
AlexaFluor™ 647	Donkey anti-goat IgG	1:1000	Invitrogen, A21447

2.3 Gene Expression Analysis

RNA was extracted and purified using the RNeasy Mini Kit (Quiagen, 74104) according to the manufacturer's instructions. Reverse transcription was carried out using 1 µg of RNA and SuperScript™ IV Reverse Transcriptase (Invitrogen, 18090050) as per the manufacturer's instruction. GoTaq qPCR Master Mix (Promega, 740955) was used for the PCR reactions and run using a 7500 Real-Time PCR System (Applied Biosystems). Relative expression was calculated using the ΔΔCt method and human *PBGD* for normalization. The primers used are listed in **Table 6**.

Table 6 - Primer Sequences Used for qRT-PCR.

All sequences are 5'→ 3'.

Primer Name	Forward	Reverse
HS2ST1	GCAGTTCCAATTCATCAGAGC	TTTGTGCGAGGATGTAGAGGT
PBGD	GGAGCCATGTCTGGTAACGG	CCACGCGAATCACTCTCATCT
NANOG	GCAGAAGGCCTCAGCACCTA	AGGTTCCCAGTCGGGTTCA
OCT4	GCTCGAGAAGGATGTGGTCC	CGTTGTGCATAGTCGCTGCT
NESTIN	CTCCAGAAACTCAAGCACC	TCCTGATTCTCCGCGGCCA
PAX6	CTTTGCTTGGGAAATCCGAG	AGCCAGGTTGCGAAGAACTC
SOX1	GGAATGGGAGGACAGGATTT	AACAGCCGGAGCAGAAGATA
B3TUB	AACGAGGCCTCTTCTCACAA	GGCCTGAAGAGATGTCCAAA
CXCR4	CACCGCATCTGGAGAACCA	GCCCATTTCCTCGGTGTAGTT
FOXA2	ACTCGCTCTCCTTCAACG	CCTGGTAGTAGGAGGTATCTGC
SOX17	AAGATGCTGGGCAAGTCGTG	GTTCAAATTCCGTGCGGTCC
GATA4	TCCCTCTTCCCTCCTCAAAT	TCAGCGTGTAAAGGCATCTG
FGFR1	CCCGTAGCTCCATATTGGACA	TTTGCCATTTTTCAACCAGCG
FGFR2	GGTGGCTGAAAAACGGGAAG	AGATGGGACCACACTTTCCATA
FGFR3	CCCAAATGGGAGCTGTCTCG	CCCGGTCCTTGTCAATGCC
FGFR4	CCATAGGGACCCCTCGAATAG	CAGCGGAACTTGACGGTGT
PBGD	GGAGCCATGTCTGGTAACGG	CCACGCGAATCACTCTCATCT

Chapter 3

Generation of a CRISPR/Cas9 Knockout of N-Sulphoglucosamine Sulphohydrolase in hiPSCs

3.1 Introduction

3.1.1 Chapter Overview

This chapter discusses the initial work carried out to generate a knockout (KO) of the gene encoding N-sulphoglucosamine sulphohydrolase (*SGSH*), an enzyme involved in the degradation of HS, in hiPSCs using CRISPR/Cas9. By creating a targeted KO via gene disruption, we intended to generate a useful tool which could be applied to better understand the mechanistic basis of disrupted HS turnover and overall function in *SGSH*$^{-/-}$ hiPSCs and the differentiated progeny. This was planned to be the initial stage of this study into the broader molecular mechanisms underlying the observed mucopolysaccharidosis type IIIA (MPS IIIA) phenotype.

3.1.2 CRISPR/Cas9 Targeting of SGSH

The *SGSH* gene is at chromosomal location 17q25.3 and contains 8 exons, and the protein consists of 560 amino acids. To date, there are 148 mutations reported within *SGSH* of which 118 are missense (Stenson *et al.*, 2003). Generation of an hiPSC knockout of *SGSH* will allow an in-depth investigation into the molecular mechanisms behind the disease and the effect of abnormal HS accumulation on hiPSC differentiation. The chosen delivery method of the CRISPR/Cas9 genome editing machinery was plasmid-based delivery, described in section 1.8.1. This strategy involved using homology directed repair (HDR) to insert an antibiotic resistance gene (puromycin and/or blasticidin) to disrupt the *SGSH* gene. Targeted cells are subjected to an antibiotic selection pressure to ensure that only the cells with the inserted resistance gene, and therefore *SGSH* knockout are isolated. Using both antibiotics will ensure that homozygous clones can be selected as some cells may incorporate different resistance cassettes into each allele (**Figure 14**). This strategy has been used previously at The University of Nottingham, to target cardiac genes in hiPSCs, with high levels of success (Mosqueira *et al.*, 2018).

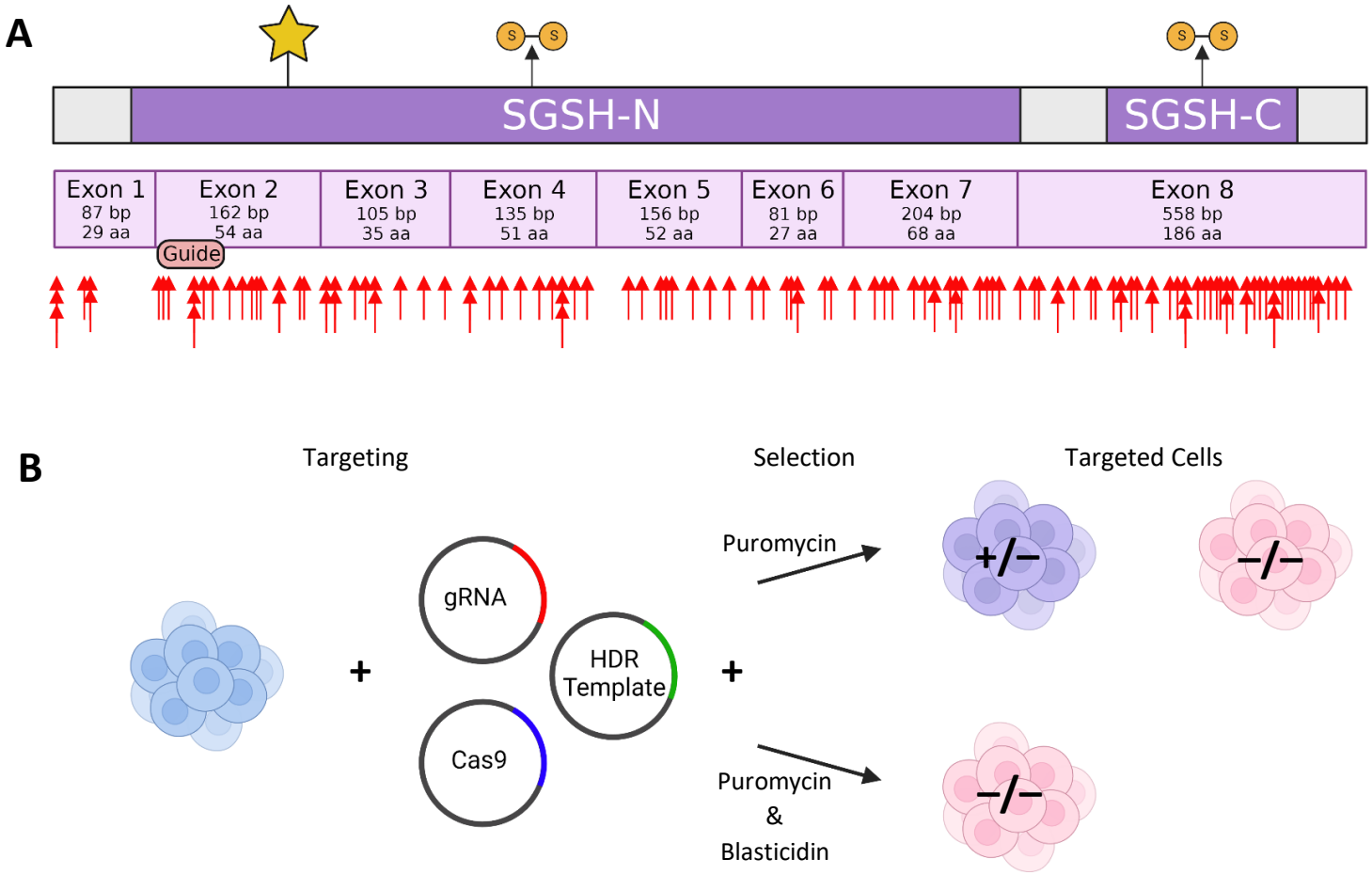


Figure 14 - Targeting Strategy to Disrupt SGSH. A) Schematic of the *SGSH* gene showing the number of base pairs (bp) and amino acids (aa) per exon, the active site (yellow star), disulphide bonds (yellow circles) which are important for the stability and conformation of the protein, CRISPR guide location, reported patient mutations (red arrows). B) HDR targeting and dual antibiotic resistance strategy ensures homozygous clones are generated by simultaneous delivery of puromycin and blasticidin HDR plasmids as different cassettes can be inserted into each allele. The hiPSCs are then transfected with both HDR plasmids, the Cas9 plasmid and the plasmid containing the gRNA.

3.2 Materials and Methods

3.2.1 gRNA Design

A plasmid-based delivery and HDR strategy was used to insert an antibiotic resistance gene in *SGSH* (plasmid schematics can be found in **Appendix 1 and 2**). Ideally, the guide would have been designed to be as close to the start site of the gene as possible but as no suitable primers could be found due to several repetitive elements, a site in exon 2 was chosen instead. The guide is upstream of the region encoding the active site (**Figure 14**). The online bioinformatics tool http://crispr.mit.edu/ was used to generate a list of guide sequences; the gRNA with the highest score and least number of off-targets was chosen, and the restriction site sequences added. Arms of homology were then chosen around the guide sequences and two primers were designed for both the left arm and right arm and the relative restriction sites were engineered onto them (**Table 7**).

Table 7 - Primer Sequences for gRNA and Arms of Homology.

All sequences 5' → 3'. Red bases indicate restriction sites (guide = BsaI, left arm forward = SalI, left arm reverse – AscI, right arm forward = BlpI, right arm reverse = NsiI).

Name	Forward	Reverse
SGSH Guide	ACCGCCTCACGCAGCGGATG ACGG	AAACCCGTCATCCGCTGCGT GAGG
Left Arm	TGCCTAGTCGACTAACACGC CCACCCAGGTAGC	CTCAAAGGCGCGCCTCATCC GCTGCGTGAGGTGG
Right Arm	ACGCAGGCTNAGCCGGAGG CTTTGAGAGTGGCG	TGTGTGATGCATACTCTTATG TGGCCCCTGGAGCA

3.2.2 Generation of Arms of Homology

To determine optimal PCR running conditions, a gradient was set up where four different annealing temperatures were tested for both the right and left arms of homology (**Table 8**). Each temperature was also tested with and without 5X Q5 GC Enhancer. Q5® High-Fidelity DNA Polymerase (NEB, M0491S) was used as

per the manufacturer's instructions. Resulting PCR products were diluted in 6X Purple Gel Loading Dye no SDS (NEB, B7025S) and separated on 1.5% agarose gels in 1X Tris-acetate-EDTA (TAE) buffer. Gels were stained using GelRed® Nucleic Acid Gel Stain (Biotium, 41003) and visualised on a UV illuminator (BioRad). The Quick-Load 1 kb ladder (NEB, N0552S) was used to identify bands of correct size.

Conditions that yielded the highest amount of the correct PCR product were used for further PCRs. Following PCR of the arms of homology, an agarose gel was run as above, the bands excised, and arms of homology purified using the NucleoSpin Gel and PCR Clean-up Kit (Machery Nagel, 740609) following the manufacturer's instructions.

Table 8 - PCR Cycling Conditions and Annealing Gradient.

Step		Temperature (°C)				Time
Initial Denaturation		98				30 seconds
Denaturation	35 Cycles	98				10 seconds
Annealing		62.8	66	68.5	70	30 seconds
Extension		72				1 minute
Final Extension		72				2 minutes
Soak		4				∞

3.2.3 Targeting Plasmid Generation

Arms of homology, SGSH guides, and plasmids were digested to create sticky ends using the relevant enzyme (**Table 9**) and CutSmart Buffer™ (NEB, B6004S) as per the manufacturer's instructions. The digestion was then run on an agarose gel and the bands excised and purified as described previously. Arms of homology and digested pUC57, or SGSH guide and linearized U6-gRNa vector, were ligated together using T4 DNA ligase (NEB, M0202S) as per the manufacturer's instructions.

Transformation of ligated plasmids into One Shot™ TOP10 Chemically Competent *E. coli* (Invitrogen, C404010) was carried out prior to cloning. Plasmids were then purified using the Wizard® Plus SV Minipreps DNA Purification System (Promega, A1330). Plasmid sequences were confirmed using Sanger sequencing (outsourced to Eurofins) and SnapGene (Insightful Science) was used for result analysis.

Table 9 - Restriction Enzyme Details.

Name	Enzyme	Plasmid	Details
5’ Right Arm	BlpI	pUC57	NEB, R0585S
3’ Right Arm	NsiI		NEB, R0127S
5’ Left Arm	SalI		NEB, R0128S
3’ Left Arm	AscI		NEB, R0558S
SGSH Guide	BsaI	U6-gRNA	NEB, R3733

3.3 Results

3.3.1 Developing Homology Directed Repair Plasmids

The main aim was to generate HDR plasmids that contain either a puromycin or blasticidin resistance cassette and the right and left arms of homology. This was done by synthesising the arms of homology and ligating them into the digested plasmids.

3.3.1.1 Determining Optimal PCR Conditions

Initially, the annealing temperatures were calculated using an online tool (https://tmcalculator.neb.com/#!/main) and the recommended temperatures were used for the PCR. However, the agarose gel showed that the left arm of homology was not successfully synthesised. Therefore, it was decided that a gradient run would be the best way to screen several temperatures to determine the optimal conditions for each arm of homology. The agarose gel showed that the right arm of homology was able to be synthesized at various annealing temperatures, both with and without GC Enhancer. In contrast, the left arm required a much more specific annealing temperature. Two conditions produced the correct PCR product (70°C and 68.5°C, both with GC enhancer), but an annealing temperature of 70°C produced a cleaner and larger band on the agarose gel indicating a higher yield of PCR product (**Figure 15A**). Therefore, an annealing temperature of 70°C with the addition of GC enhancer was taken forward as the optimal PCR conditions for both the right and left arms of homology. It is important to note that the same amount of product was loaded into each well of the gel. Following this, both arms were synthesised at 70°C with GC Enhancer which yielded clean bands of the correct size (**Figure 15B**).

3.3.1.2 Digesting Arms of Homology

Initially, there were several issues with the digestion – gels appeared smeared or multiple products were separated on the gel. After re-synthesizing the arms of homology and using fresh enzymes and CutSmart buffer, the arms were digested as expected. However, the left arm digestion appears to have generated a much cleaner product when compared with the right arm (**Figure 15C**).

3.3.1.3 Ligating Arms of Homology

Following ligation of the PCR products into the plasmids, they were then transformed into *E. coli*. This was undertaken with varying results – two times out of six, no *E. coli* colonies were present, indicating that the plasmids had not been incorporated and the bacteria could therefore not survive under antibiotic selection pressure. Where colonies formed and plasmids were purified, Sanger sequencing revealed the arms of homology had not been incorporated into the plasmids.

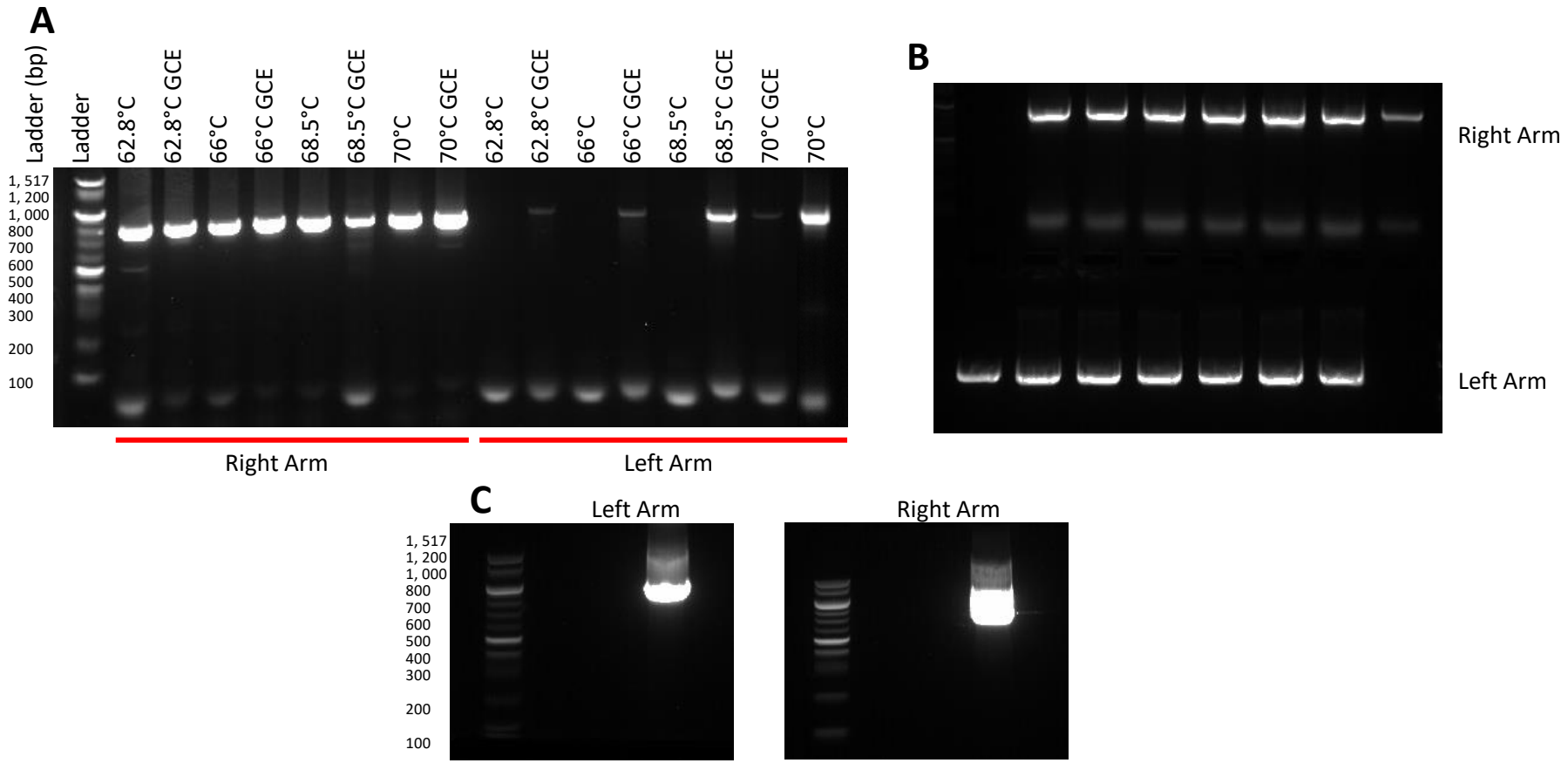


Figure 15 – Agarose Gels of Arms of Homology. A) Gradient PCR to determine optimal annealing temperatures (GCE = GC Enhancer added). Optimal conditions for both the left and right arm were 70°C annealing temperature with GC enhancer. B) Arms of homology using chosen PCR conditions showing clean bands. C) Arms of homology following digestion with relevant enzymes

3.4 Discussion

The work discussed in this chapter was designed to establish an hiPSC model to study MPS IIIA by the generation of a CRISPR/Cas9 knockout of the *SGSH* gene. At the start of this project there were no reports of such a model and most of the literature continues to focus on patient-derived iPSC models such as those discussed in Chapter 1.

3.4.1 Is Plasmid Delivery the Correct Choice For CRISPR/Cas9 Knockout of SGSH?

Normal practice is for the start site of a protein coding region to be targeted when designing a KO strategy. However, due to the rich GC content and abundance of repetitive elements in Exon 1 of *SGSH*, PCR amplification of the arms of homology would have been a challenge, so exon 2 was selected as it was still possible to target the location of the active site. However, the generation of the plasmids, involving digestion and subsequent ligation proved very challenging. Once the issues with these steps had been resolved, incorporation of the plasmids into *E. coli* failed or, where successful, the plasmids did not contain the correct targeting sequences. Plasmid based delivery was chosen as it was routinely used at the University of Nottingham with success (Mosqueira *et al.*, 2018). However, this method requires a considerable number of extra steps to generate the plasmids which, in this study, unfortunately failed.

The use of ribonucleoprotein (RNP) delivery of the Cas9 and gRNA would not only eliminate the need for transcription and translation intracellularly, but also the need for plasmid generation. RNP delivery is a much faster method of CRISPR/Cas9 delivery and reduces off target effects while maintaining a high gene editing efficiency (Zhang *et al.*, 2020). In this study, plasmid generation was the primary issue at all stages – digestion, ligation, and *E. coli* transformation – so circumventing this stage of the gene editing process would be preferential and could be achieved using RNP delivery.

3.4.2 Future Work

Although this work was halted at a very early stage, there is still a need for the detailed comparison of CRISPR/Cas9 engineered hiPSC lines compared with patient derived cell lines. The generation of such a model would provide a method of investigating *SGSH* mutations of different developmental pathways and also understand the effect of abnormal GAG degradation on GAG biosynthesis. Furthermore, hiPSCs could be targeted at varying locations using HDR to introduce very specific point mutations. This would allow for the creation of a library of *SGSH* mutated hiPSC lines which could all be compared against each other. The generation of such a library could provide an understanding of the *SGSH* gene and protein and determine regions of the gene that are responsible for differing severities of MPS, and why. Additionally, a CRISPR/Cas9 hiPSC model would not require collection of material from patients and therefore not require ethics approval, so work could begin immediately. Once patient material is able to be sourced, a comparison between CRISPR/Cas9 hiPSC lines and patient derived hiPSC lines could be carried out to determine whether significant difference exist. This may give an insight into whether CRISPR/Cas9 models are actually a suitable method for the investigation of GAG related diseases. Generation of a patient-derived hiPSC line was initially part of the plan for this study and would have resulted in a fully characterized CRISPR/Cas9 derived and patient derived hiPSC line with *SGSH* alterations. This would have allowed for novel comparison of the differentiation capabilities and disease influence between the two differently derived cell lines and appropriate controls.

In the future, RNP delivery of the Cas9 and gRNA would be recommended due to the issues with the plasmid vector generation and *E. coli* transformation. Furthermore, RNP delivery would reduce off target effects and allow for faster genome targeting. Directed differentiation of the hiPSC lines for a longer time would be recommended to see the effects of the accumulation of abnormal HS. It can be hypothesized that significant differences would occur with the differentiations, particularly with the neural lineage due to the abnormal build-up of monosialic gangliosides (as a result of

partially degraded HS accumulation) which are present at higher levels and have a range of functions in the neural tissues.

Chapter 4

Generation of a Heparan Sulphate 2-O-Sulphotransferase CRISPR/Cas9 Knockout in hiPSCs

4.1 Introduction

4.1.1 Chapter Overview

The focus of this chapter is to better understand the impact of loss of heparan sulphate 2-O-sulphotransferase (HS2ST1) activity. A collaboration with the Kutsche Lab at the Institute of Human Genetics in Hamburg provided access to heparan sulphate (HS) from patients who have novel mutations in the gene encoding HS2ST1. Material was collected from fibroblast cultures derived from these patients and disaccharide analysis was carried out at the University of Nottingham.

In a subsequent study, presented in this doctoral thesis, we carried out CRISPR/Cas9 targeting of *HS2ST1* in hiPSCs and characterization of the impact on HS composition in these cell lines.

4.1.2 CRISPR/Cas9 Targeting of HS2ST1

The *HS2ST1* gene is at chromosomal location 1p22.3 and contains 7 exons encoding a protein of 357 amino acids. Several amino acids within the protein are key for the essential function of the HS2ST1 enzyme. For example, two histidine residues at amino acid positions 140 and 142 (shown as the active site in **Figure 5 and 18**) are essential for the catalytic activity of the enzyme. Furthermore, a lysine, serine and two threonine residues at position 83, 86, 87 and 84, respectively, are all essential for 3'-phosphoadenosine 5'-phosphosulphate (PAPS) binding and loss of these results in reduced HS2ST1 activity. PAPS provides the sulphate group which HS2ST1 transfers to the HS chain. Amino acid 80 is an arginine and is essential for substrate binding. Additionally, arginine (178) and asparagine (181) are located close to the PAPS binding site in the final protein conformation and loss of these amino acids results in reduced HS2ST1 activity. It was reported that these two amino acids were essential for not only substrate binding, but also stabilization of PAPS binding. These

findings indicate that several areas along the *HS2ST1* gene are essential for the correct function of the enzyme.

To date, four mutations have been reported in patients (**Table 1, Figure 5**). Generation of an hiPSC knockout of *HS2ST1* would allow for an in-depth investigation into the molecular mechanisms behind the novel disease described in this chapter. Additionally, as HS2ST1 results in developmental defects, an investigation into the affected tissues can be carried out using directed differentiation of the KO hiPSCs to different germ layers and cell types. Furthermore, KO hiPSCs can be used to assess differences in fundamental signalling pathways to determine how loss of HS2ST1 may affect development. Ribonucleoprotein (RNP) delivery of the CRISPR/Cas9 machinery was used, combined with non-homologous end joining (NHEJ) to repair the DNA (**Figure 12 and 13**).

4.2 Materials and Methods

4.2.1 Fibroblast Cell Culture

Fibroblast isolation and culture was carried out by the Kutsche Lab. Primary fibroblasts were obtained from a skin biopsy from three patients and four healthy individuals (control subjects). They were cultured in Dulbecco's Modified Eagle Medium (DMEM; ThermoFisher) supplemented with 10% Foetal Bovine Serum (FBS; GE Healthcare) and Penicillin-Streptomycin (Pen-Strep; 100 U/mL and 100 mg/mL, respectively; ThermoFisher).

4.2.2 Extraction of Glycosaminoglycans from Conditioned Medium and Cell Lysates

Cultured primary fibroblasts were seeded in T75 flasks while hiPSCs were seeded into T25 flasks and conditioned media was collected after every media change. After 3 days, cells were washed with 1 X DPBS, then lysed using 1% (vol/vol) Triton X-100/PBS for 3 hours at room temperature while shaking to prepare a cell/matrix extract. Prepared conditioned media or cell extract samples were treated with 100 mg/mL Pronase (Roche Diagnostics, 0165921001) for 4 hours at 37°C. Cell extracts were then additionally incubated for 15 min at 90°C to deactivate the Pronase, before treating with 14 mg/mL DNase I (Sigma, DN25) and in 10 mM $MgCl_2$ (Sigma, M8266) overnight at 37°C, to prevent DNA interference with downstream processing.

All preparations were loaded onto pre-equilibrated 1 mL diethylaminoethyl (DEAE)-Sephacel anion exchange columns (Sigma, I6505) equilibrated in 1.5 M NaCl, 20 mM $NaH_2PO_4.H_2O$ (pH 7.0) and washed with 50 column volumes (cv) of 0.25 M NaCl, 20 mM $NaH_2PO_4.H_2O$ (pH 7.0) to remove non-binding material, including hyaluronan. Sulphated GAGs were then eluted with 5 x cv of 1.5 M NaCl, 20 mM $NaH_2PO_4.H_2O$ (pH 7.0) and subsequently desalted using PD-10 Sephadex G-25M pre-packed columns (GE Healthcare, 17-0851-01), eluting the purified GAGs with H_20.

4.2.3 Purification of AMAC Fluorescent Label

Prior to labelling of HS disaccharides, 2-aminoacridone (AMAC; Sigma, 06627) required purification by hydrophobic interaction chromatography. AMAC was solubilised in 85% dimethyl sulfoxide (DMSO; Sigma, S-002-D)/15% acetic acid to give a 10 mg/mL stock and diluted 1:4 in H_20. 1 mL was loaded onto a pre-

equilibrated 2 mL Macro-prep t-butyl column (BioRad, 64-17-5). 20 mL of 20% methanol was added to elute the impurity. AMAC was then eluted using 20 mL of 100% methanol. Purified AMAC was lyophilised and reconstituted in 85% DMSO/15% acetic acid to a final concentration of 0.1 M.

4.2.4 Preparation of AMAC-labelled Disaccharides

In order to detect HS, the purified GAGs need to be digested and fluorescently labelled. Purified GAGs (from section 4.2.2) were lyophilised, then resuspended in 100 µL of 0.1 M sodium acetate and 0.1 mM calcium acetate (pH 7.0) and digested using 0.8 mIU each of heparinase I, II, and III (*Flavobacterium heparinum*, Iduron) for 12 hours at room temperature. The resulting digests were again lyophilised, then re-suspended in 10 µL of 0.1 M AMAC and incubated at room temperature for 20 minutes. Following this, 10 µL of 1 M $NaBH_3CN$ (Sigma, 156159) was added, and incubation continued at room temperature overnight in the dark. Samples were diluted in 80 µL H_20 prior to analysis.

The AMAC labelled samples were separated by reverse phase high performance liquid chromatography (RP-HPLC) in duplicate using a Zorbax Eclipse XDB-C18 RP-HPLC column (35 µM, 2.1 mm x 150 mm; Agilent Technologies). The flow rate was 200 µL/min and injection volumes were 20 µL for samples and 5 µL for standards. The elution solvents were 20 mM ammonium acetate (solvent A) and acetonitrile (solvent B). A linear gradient of 95:5 (0 mins), 95:5 (5 mins), 80:20 (30 mins), 0:100 (35 mins), 95:5 (55 mins), 95:5 (65 mins) was used (A:B). Disaccharides were detected by fluorescence (excitation 425 nm, emission 520 nm) and were identified and quantified in comparison with known amounts of commercial disaccharide standards (Iduron). Peak integration was carried out and a pre-determined labelling efficiency factor (section 4.2.5) was applied for disaccharide quantification.

4.2.5 Preparation and Analysis of Disaccharide Standards For RP-HPLC

Seven HS disaccharide standards (Iduron, HD001, 2 and 4 to 8) were used for analysis. For each standard, aliquots from 1 mg/mL stocks were freeze-dried and AMAC labelled as above before being diluted in H_20 to give 50 µg/mL stocks. For RP-HPLC analysis, all seven standards were mixed together to give a final concentration

of 5 µg/mL for each which corresponded to 23 ng of each standard per run. Disaccharide labelling efficiencies were calculated for all standards by running each standard separately on the HPLC. The corresponding UV and fluorescence peak areas for each standard were obtained and labelling efficiency calculated using the equation below:

$$Labelling\ Efficiency = \frac{UV\ peak\ area}{Fluorecence\ peak\ area}$$

4.2.6 Chinese Hamster Ovary Cell Culture

Chinese Hamster Ovary (CHO) cells (wild type K1 cells and chemically mutated F17 cells (lacking 2-O-sulphated HS), kindly provided by Professor Jeffrey Esko (Bai and Esko, 1996)) were routinely cultured in Ham's F12 Nutrient Mix (ThermoFisher, 11765054) supplemented with 7.5% (v/v) FBS (Life Technologies, 10695023) and 100 µg/mL Pen-Strep (Gibco, 15140-122). The medium was changed every 48 hours. Once 70% confluent, CHO cells were passaged by incubating cells with trypsin/EDTA (Life Technologies, 25200056) for 2 minutes at 37°C before collecting cells in CHO culture medium. Cells were pelleted at 300 x g for 3 minutes and resuspended in culture medium before seeding at approximately 20,000 cells/cm^2. Cells were cultured at 37°C, in 5% CO_2 in a humid incubator.

4.2.7 crRNA Design

RNP delivery of the Cas9 and gRNA was used to target *HS2ST1*, combined with a NHEJ approach for DNA repair. To improve the potential for creation of a knockout, *HS2ST1* was targeted in two different locations, Exon 2 and Exon 3 so two different crRNAs were designed (**Table 10**). Guides were checked using a CRISPR/Cas9 guide RNA design checker (Integrated DNA Technologies) which also provided the PAM sequence, on-target and off-target score.

Table 10 - crRNa Sequences

Target Location	Sequence
Exon 2	GACATACAATGGATGGCCCT
Exon 3	TCATGGACACGTTTCTTACT

4.2.8 gRNA Duplex and RNP Complex Formation

Prior to RNP complex formation, Alt-R™ CRISPR/Cas9 tracrRNA (Integrated DNA Technologies, 1072533) and Alt-R™ CRISPR-Cas9 crRNA (Integrated DNA Technologies, custom sequence, **Table 10**) were resuspended in IDTE buffer (10 mM Tris, 0.1 mM EDTA, pH 8.0; Integrated DNA Technologies, 11-05-01-13) to a final concentration of 200 µM. gRNA duplex formation was achieved by mixing equal concentrations of crRNA and tracrRNA to a final duplex concentration of 100 µM before incubating at 95°C for 5 minutes. The RNP complex was formed by combining 120 pmol of gRNA duplex with 104 pmol of Edit-R Cas9 Nuclease Protein NLS (Horizon Discovery, CAS11730) and 1X DPBS up to a final volume of 5 µL and incubating at RT for 20 minutes.

4.2.9 Transfection and Picking

All cells used for gene targeting were cultured in E8 medium on Matrigel® (Corning, 356234) coated plates. For the coating, Matrigel® was diluted in cold Dulbecco's' Modified Eagle Medium/Nutrient Mixture F12 (DMEM/F12, ThermoFisher, 1132003) and set overnight at 4°C or for at least 1 hour at 37°C.

Cells were dissociated with Accutase solution (Sigma, A6964) for 5 minutes at 37°C and collected in E8 medium supplemented with 10 µM ROCKi. For each nucleofection, 200,000 cells were combined with the RNP complex and Alt-R™ Cas9 Electroporation Enhancer (Integrated DNA Technologies, 1075915). Three different nucleofection conditions were used – exon 2 targeted cells, exon 3 targeted cells and cells where both exons 2 and 3 were targeted together. The P3 Primary Cell 4D-Nucleofector™ X Kit (Lonza, V4XP-3032) was used for nucleofection on a 4D-Nucleofector™ Unit using program CA-137, as per the manufacturer's instructions.

Following nucleofection, cells were immediately transferred to a 37°C incubator for 5 minutes before transferring into a pre-prepared Matrigel® coated 24 well plate in fresh E8 medium supplemented with 10 µM ROCKi and 0.5% Pen-Strep. After 24 hours, medium was replaced with fresh E8 medium without ROCKi. Once hiPSCs were confluent, they were dissociated to single cells with accutase solution for 5 minutes at 37°C, collected in fresh

medium and pelleted at 300 x g for 3 minutes. hiPSCs were resuspended in cloning medium – E8 medium supplemented with 10% CloneR™ (Stem Cell, 05889) and 0.5% Pen-Step – and seeded on Matrigel® at 400 and 600 cells per 35 mm dish (two separate seeding densities were used). After 48 hours, medium was replaced with fresh cloning medium. 24 hours later (72 hours after seeding) fresh cloning medium was added at 25% of the initial seeding volume. 96 hours after initial seeding, medium was replaced with fresh E8 medium with no supplements and changed daily hereafter.

Once colonies had formed, clones were picked using a Stem Cell Cutting Tool (Vitrolife, 14601) and seeded into 48 well Matrigel® coated plates in E8 medium supplemented with 10 μM ROCKi which was changed to fresh E8 medium daily. hiPSC clones were gradually expanded into 6 well plates and once confluent, were frozen in E8 medium, 10 μM ROCKi and 10% DMSO and their DNA harvested for sequencing. Prior to knockout confirmation, hiPSCs were thawed on Matrigel® before being passaged onto VTN-N and cultured for at least two passages to ensure removal of exogenous GAG (from the Matrigel®).

4.2.10 Genotyping

DNA was extracted from hiPSC clones using a DNeasy Blood & Tissue Kit (Qiagen, 69556) as per the manufacturer's instructions. 50 ng of DNA was amplified by GoTaq G2 DNA Polymerase (Promega, M7845) in the PCR reaction which used primers designed to capture the Cas9 cut site (**Table 11**). PCR cycling conditions can be found in **Table 12**. PCR products were diluted in 6X Purple Gel Loading Dye No SDS (NEB, B7025S) and separated on a 1% agarose gel, stained with GelRed® Nucleic Acid Gel Stain (Biotium, 41003), in 1X TAE buffer. The gel was visualized on a UV illuminator (BioRad) and a Quick-Load 1 kb ladder (NEB, N0552S) was used to confirm the correct PCR product. The remaining PCR product was cleaned up using a NucleoSpin Gel and PCR Clean-up Kit (Machery-Nagel, 740609) following the manufacturer's instructions. Sequencing of PCR products was outsourced to Eurofins and SnapGene (Insightful Science) was used for analysis.

Table 11 - Genotyping Primer Sequences

Target Location	Forward	Reverse
Exon 2	ATGGTCCAAAGTTCAGTGTTCA	GCTTGCCGTTTTGGGAACTC
Exon 3	TTCTCCCTTTTTAGGTGCGCT	TGGCAAAGGCAAGACAGTACA

Table 12 - PCR Cycling Conditions

Step		Exon 2		Exon 3	
		Temp (°C)	Time	Temp (°C)	Time
Initial Denaturation		95	2 minutes	95	2 minutes
Denaturation	35 Cycles	95	30 seconds	95	30 seconds
Annealing		59.2	20 seconds	61.2	20 seconds
Extension		72	1 minute	72	1 minute
Final Extension		72	5 minutes	72	5 minutes
Soak		4	∞	4	∞

4.2.11 Flow Cytometry

To generate a single cell suspension, hiPSCs and CHO cells were harvested in Gentle Cell Dissociation Reagent (GCDR; Stem Cell Technologies, 100-0485) for 10 minutes at 37°C and washed in relevant cell culture medium. Cells were then pelleted at 300 x g for 5 minutes and resuspended in Fluorescence-Activated Cell Sorting (FACS) buffer – 1% Bovine Serum Albumin (BSA; Sigma, 232-936-2) in 1X DPBS – to a final concentration of 1 x 10^7 cells/mL. 100 µL of cell suspension per condition was transferred into a V bottom 96 well plate for staining. All incubations were carried out for 1 hour at 4°C and all centrifugation steps were performed at 300 x g for 3 minutes at 4°C. The IgM primary antibody (10E4) was diluted in FACS buffer (**Table 13**), added to the cells and incubated for 1 hour at 4°C before centrifuging and washing three times in FACS buffer, pelleting cells between each wash. For the ScFv antibody (AO4B08), cells were incubated with primary antibody diluted in FACS buffer (**Table 13**), pelleted and washed as above before the secondary antibody (Anti-VSV-G) was

added. Secondary antibody was diluted in FACS buffer (**Table 13**) and incubated with cells before washing the cells as previously described. Cells were then incubated with the relevant fluorescent antibodies – AlexaFluor™ 488-tagged goat anti-mouse IgM (Invitrogen, 21042) and AlexaFluor™ 488-tagged goat anti-mouse IgG (Invitrogen, A11029) which were diluted 1:500 in FACS buffer and added to the cells for incubation and subsequent washing. Cells were then resuspended in FACS buffer and transferred to an ID7000™ Spectral Cell Analyzer (Sony) at The University of Nottingham Flow Cytometry Facility for analysis. A total number of 20,000 events was recorded (in the single cell gates for each sample). Resulting data was analysed using Kaluza software.

Table 13 - Antibodies Used for Knockout Confirmation via Flow Cytometry

Antibody	Isotype	Dilution	Supplier Details
10E4	Mouse IgM	1:200	Amsbio, 370255-1
AO4B08	ScFv	1:10	Gift from Toin H van Kuppevelt (Kurup *et al.*, 2007)
Anti-VSV-G	Mouse IgG1	1:1000	Sigma, SAB4200695
AlexaFluor™ 488	Goat anti-mouse IgM	1:1000	Invitrogen, A21042
AlexaFluor™ 488	Goat anti-mouse IgG	1:1000	Invitrogen, A11029

4.3 Results

4.3.1 The Heparan Sulphate Disaccharide Composition from Patients with Defects in *HS2ST1* Is Significantly Different When Compared to Control Cell Lines

As the patients with identified mutations in *HS2ST1* were the first to be reported, it was important to understand the impact on the HS structure. We would usually hypothesize that changes in 2-O-sulphation (2S) would be present due to lack of the enzyme, but it has previously been suggested that humans could not survive without this HS modification. The disaccharide composition of HS isolated from fibroblasts of one patient with *HS2ST1* genetic abnormalities and two control patients with normal genotyping was analysed to determine whether the patient showed a shift in HS sulphation patterning. AMAC labelling followed by separation and identification of disaccharides by RP-HPLC revealed a significant reduction of overall 2S in both cell lysate and conditioned medium when compared with the controls (**Figure 16A-16D**). Specifically, the total level of 2S was 0.6% in patient cell lysate compared to 15% in both control cell lysates (**Figure 16C**). Similar differences were observed in conditioned medium (patient conditioned medium 2%; control 1 11%; control 2 8%; **Figure 16D**). The trace levels of 2S seen in conditioned medium of the patient cells was likely due to HS present in the serum (serum-containing medium not exposed to cells was found to have 6% 2S; **Figure 17**). Unfortunately, fibroblasts could not survive without serum so the effects of exogenous GAG could not be mitigated. An increase in 6-O-sulphation (6S) and N-sulphation (NS) was observed in the HS from the patient cell lysate with 19% 6-O-sulphation and 66% N-sulphation compared to 12% 6S in both controls and 45% and 40% NS in control 1 and 2, respectively (**Figure 16C**). A similar pattern of altered sulphation was seen in HS isolated from the conditioned medium (**Figure 16D**). Although the increase in total 6S is not significant, the increase in the disulphated disaccharide UA-GlcNS(6S) was significantly increased in the patient medium and cell lysates (10-15% across both controls and both sample preparations). This suggests that the increase in 6S is enhanced in the N-sulphated domains with the monosulphated disaccharide UA-GlcNS was also significantly increased in the patient compared to controls (**Figures 16A and 16B**). HS was isolated from two other patients, but unfortunately, it was not possible to extract sufficient HS for analysis.

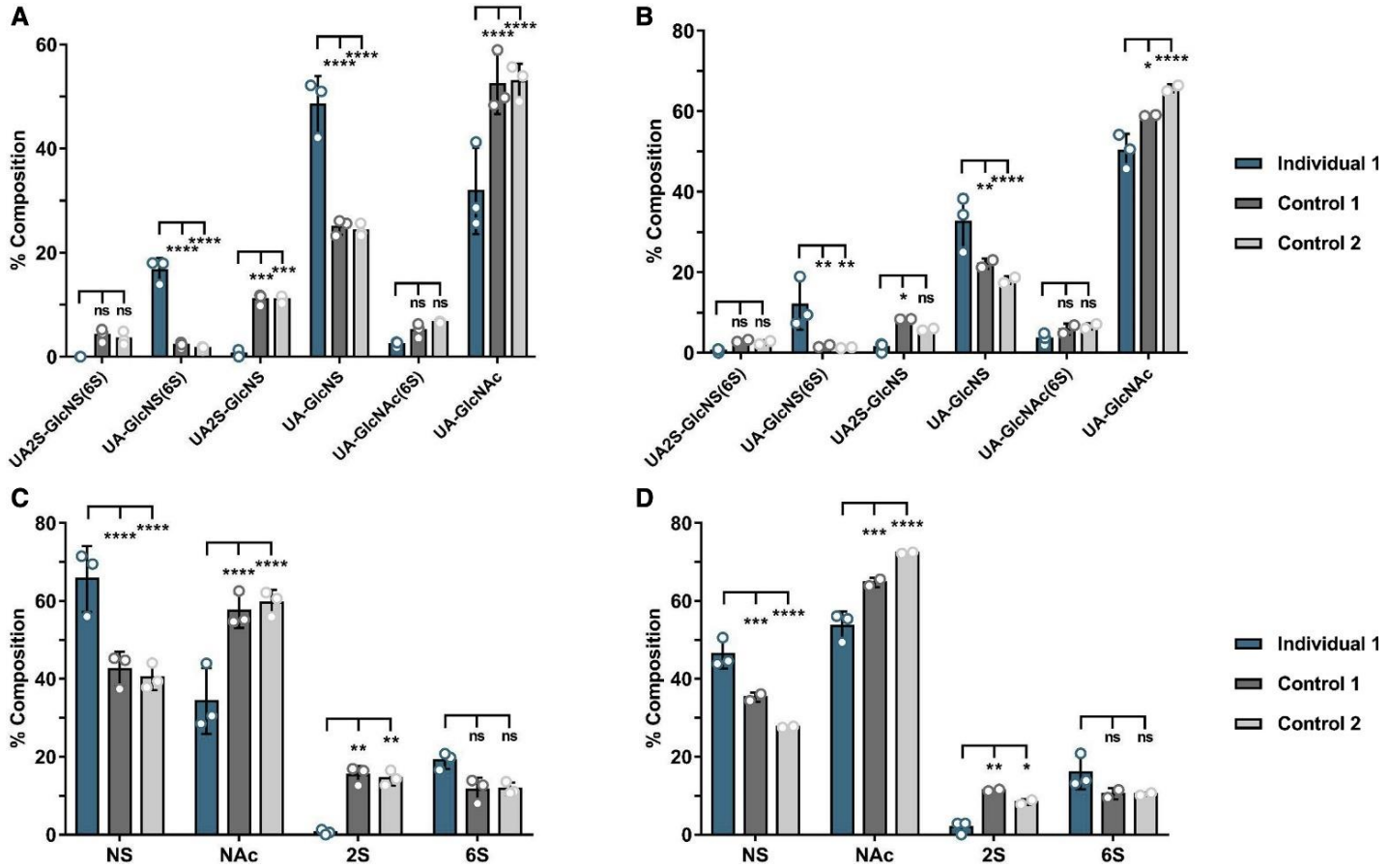


Figure 16 – Heparan Sulphate Disaccharide Composition of Patient 1. A) HS disaccharide composition of cell lysates and B) conditioned media. C) HS modification changes in cell lysates and D) conditioned media. Error bars represent ± SD from patient 1 conditioned media, patient 1 cell lysates, control 1 cell lysates and control 2 cell lysates n=3; control 1 conditioned media and control 2 conditioned media n=2. Two-way ANOVA with Tukey correction was used for statistical analysis: $^{*}p \leq 0.05$; $^{**}p \leq 0.01$; $^{***}p \leq 0.001$; $^{****}p \leq 0.0001$; ns, not significant.

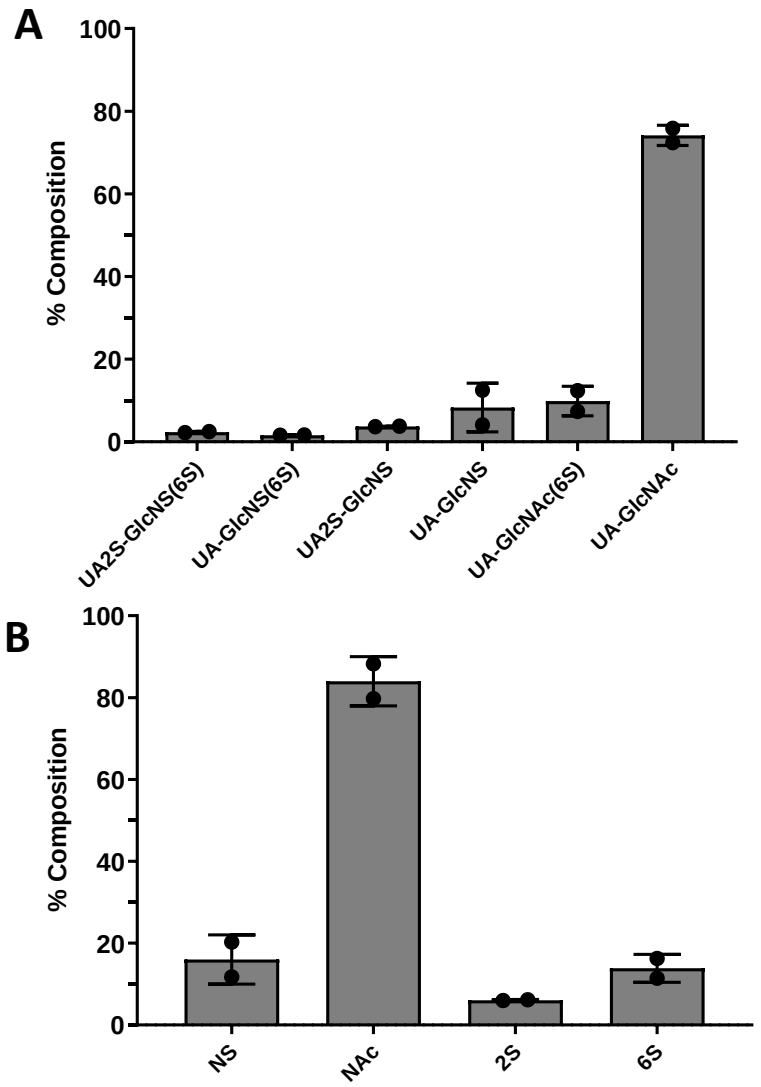


Figure 17 – Heparan Sulphate Disaccharide Composition of Serum Containing Medium. A) Detailed HS disaccharide composition. B) Overall HS modification changes. Patient fibroblasts could not be cultured in serum free medium, so the HS composition of the medium was determined for control purposes. Error bars represent ± SD from n=2

4.3.2 HS2ST1 Targeting in REBLPAT hiPSCs

The aim was to generate homozygous and heterozygous *HS2ST1* knockouts. RNP delivery and NHEJ were the methods chosen to generate knockouts due to the many benefits associated with this approach, as discussed in section 1.8.1. Furthermore, RNP CRISPR has been used at the University of Nottingham with significant success (Sara Pijuan-Galitó and Alexander Kondrashov, personal communications). **Figure 18A** shows a schematic of the *HS2ST1* gene and the two targeting sites. Two guides were designed to improve the chances of generating a knockout. Guide 1 was designed over the start site of the sulphotransferase domain and Guide 2 was designed to target the amino acids critical for the formation of the active site (Phe-Tyr-His-Gly-His). hiPSCs were targeted with each of the guides separately as well as together, generating three different targeting conditions. Unlike plasmid-based delivery and insertion of an antibiotic resistance gene, no selection pressure was used. 192 clones were generated with each one requiring sequencing. However, only 96 were able to be sequenced during the project.

Many clones showed no genetic alterations when sequences (data not shown) and so were discarded. Clones that were genetically altered had clean, single sequences, with traces matching that of REBLPAT WT hiPSCs, until the Cas9 cut site where two traces appeared superimposed (**Figure 18B**). Each clone that was picked, frozen and showed subsequent genetic alteration was provided a letter. This chapter will discuss three different clones – A, B and C which were all targeted in exon 2 only. One clone out of 96 sequences (Clone C) showed a clean sequence trace (i.e. no overlapping traces at Cas9 cut site) but had a homozygous 7 base pair deletion, which resulted in a stop codon, when compared with the REBLPAT WT sequence (**Figure 18C**). 29 clones out of 96 that were sequenced showed genetic alterations. It was not possible to infer if these resulted in enzyme knockout from sequencing alone. Furthermore, all clones appeared viable and showed no obvious differences in attachment, rate of cell death, proliferation or cell shape when assessed by eye (**Figure 18D**). Therefore, phenotypic confirmation was required to determine whether the genetic abnormalities resulted in loss of enzymatic function. As at the time of phenotyping the HS2ST1 function was unknown, each hiPSC line will be referred to by the letter it was assigned (A, B or C).

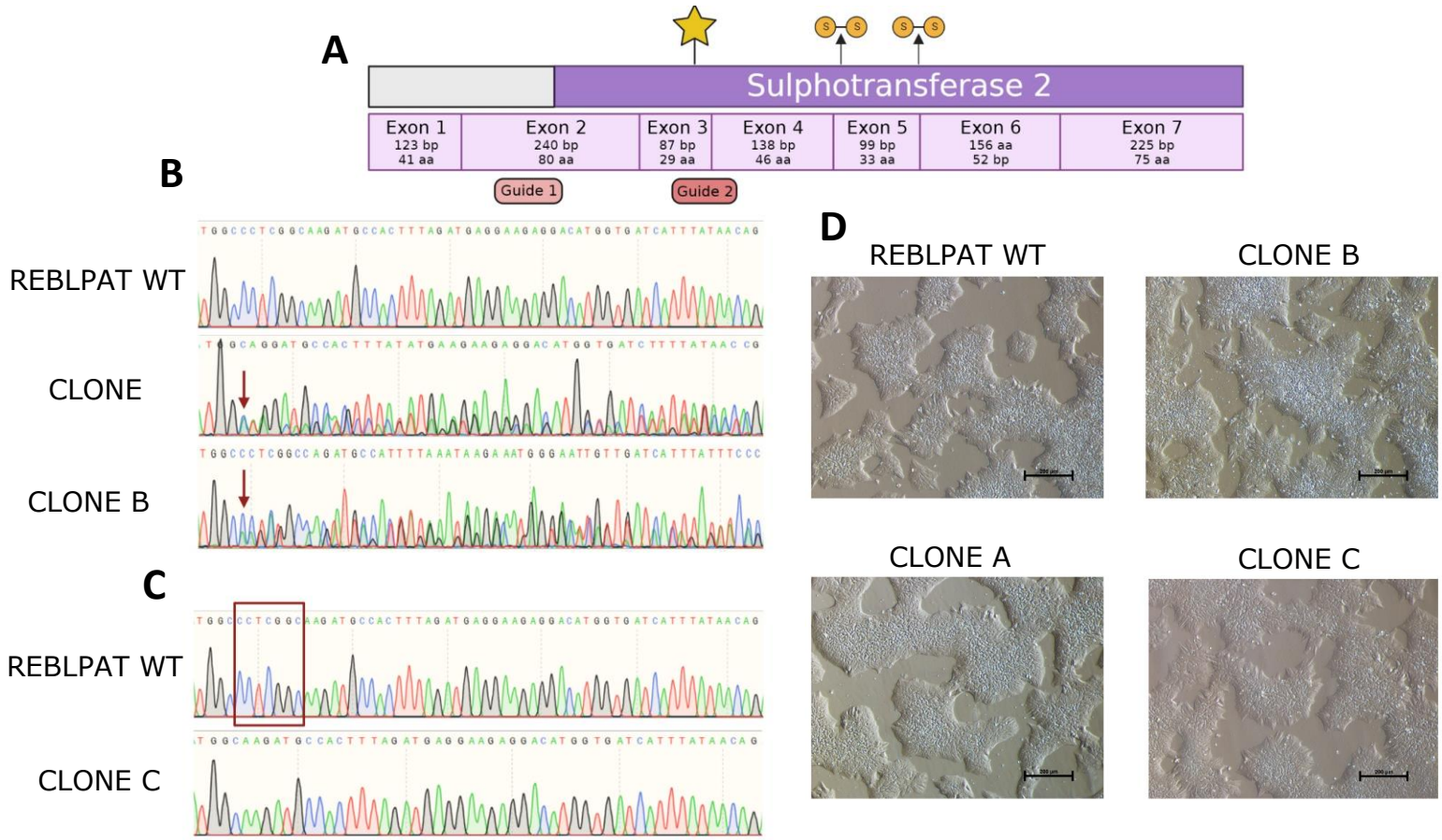


Figure 18 - CRISPR Targeting of HS2ST1. A) Schematic of the HS2ST1 gene showing the active site (yellow star), disulphide bonds (yellow circles) which are important for the stability and conformation of the protein, exon number, base pair (bp) number and amino acid (aa) number and CRISPR guide locations. B) Sequencing showing cells targeted by Cas9 (at site of red arrow) resulting in 2 overlapping traces C) Sequencing showing a 7 base pair knockout (red square). D) Light microscope images of targeted cells show knockout genetic alteration does not affect morphology. Scale bars 200 µm

4.3.3 HS2ST1 mRNA Is Still Present in All Clones

When using the HDR strategy, as in Chapter 3, qRT-PCR can be an effective way to determine knockout as the mRNA transcription of the gene is completely disrupted by the insertion of the antibiotic resistance cassette. For this reason, qRT-PCR using HS2ST1 primers was used as a potential knockout confirmation method in the present study despite the CRISPR targeting strategy using NHEJ. Over three repeats, the mean change in HS2ST1 expression was 0.98-, 1.02- and 1.14-fold in Clone A, Clone B and Clone C respectively when compared to REBLPAT WT cells (**Figure 19**). These changes were not significant.

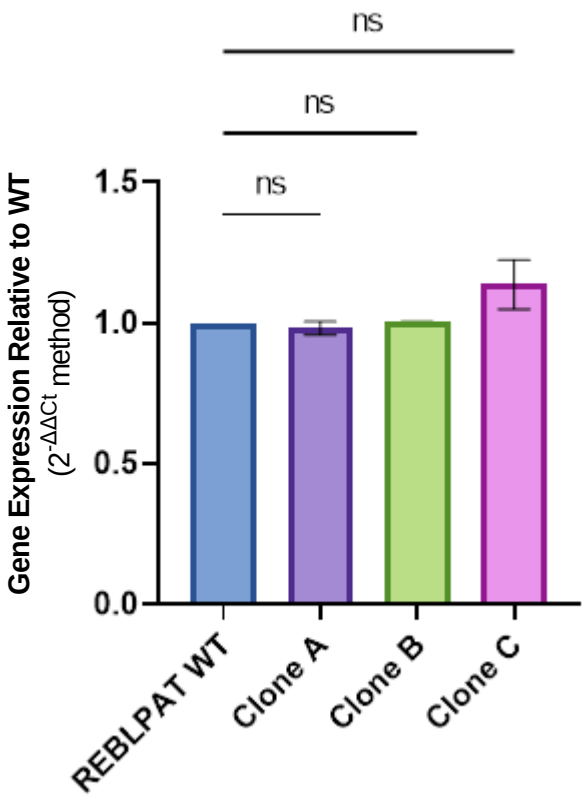


Figure 19 - HS2ST1 mRNA Expression. qRT-PCR was used to compare HS2ST1 gene expression in Clone A, Clone B and Clone C to REBLPAT WT hiPSCs to determine HS2ST1 knockout. HS2ST1 expression is not significantly different in any hiPSC clone. Error bars represent ± SEM from n=3 for REBLPAT WT, Clone A and Clone C, n=3 for Clone B. Tukey's multiple comparisons test was carried out to determine statistical differences.

4.3.4 Heparan Sulphate 2-O-Sulphation is Undetectable in Knockout Cell Lines

An antibody for endogenous levels of the HS2ST1 enzyme is not available, so we used instead a well validated antibody that detects the 2S modification generated by the enzyme (AO4B08). This antibody was previously used to characterize HS from the *HS2ST1*$^{-/-}$ mouse (Kurup *et al.*, 2007). 10E4, which recognises an N-sulphated motif common within HS, was also used to confirm the presence of HS (David *et al.*, 1992). Therefore, loss of AO4B08 reactivity, when 10E4 staining was present could be attributed to specific loss of 2S and not HS in general. Three control cell lines were used – CHO *HS2ST1*$^{-/-}$ (to help determine the optimal antibody concentration), CHO WT and REBLPAT WT hiPSCs. All cell lines stained positive for 10E4, via immunocytochemistry (**Figure 20 and 21**). This was confirmed by flow cytometry, though peaks for Clone B and Clone C appear to show differential binding of the 10E4 antibody to the HS (**Figure 22**). It is important to note that staining appears concentrated around the edge of hiPSC colonies. Several different cell lines have been repeatedly imaged on different microscopes, but the edging-effect is still seen. Furthermore, other people have also reported seeing this effect, yet the reasons for its appearance remain unknown (Jamie Thompson, personal communication).

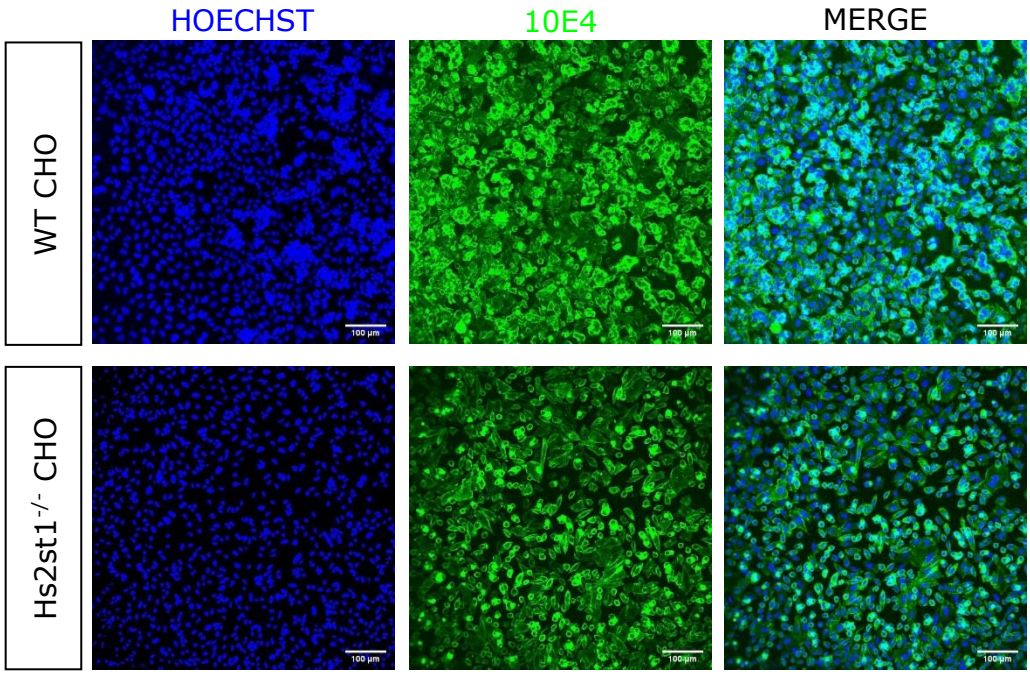


Figure 20 – Hs2st1$^{-/-}$ CHO Cell Lines Produce Heparan Sulphate. Immunofluorescent images showing expression of 10E4 (HS) in WT and *Hs2st1*$^{-/-}$ CHO cell lines. HS expression appears similar in both lines. Scale bar 100 μm

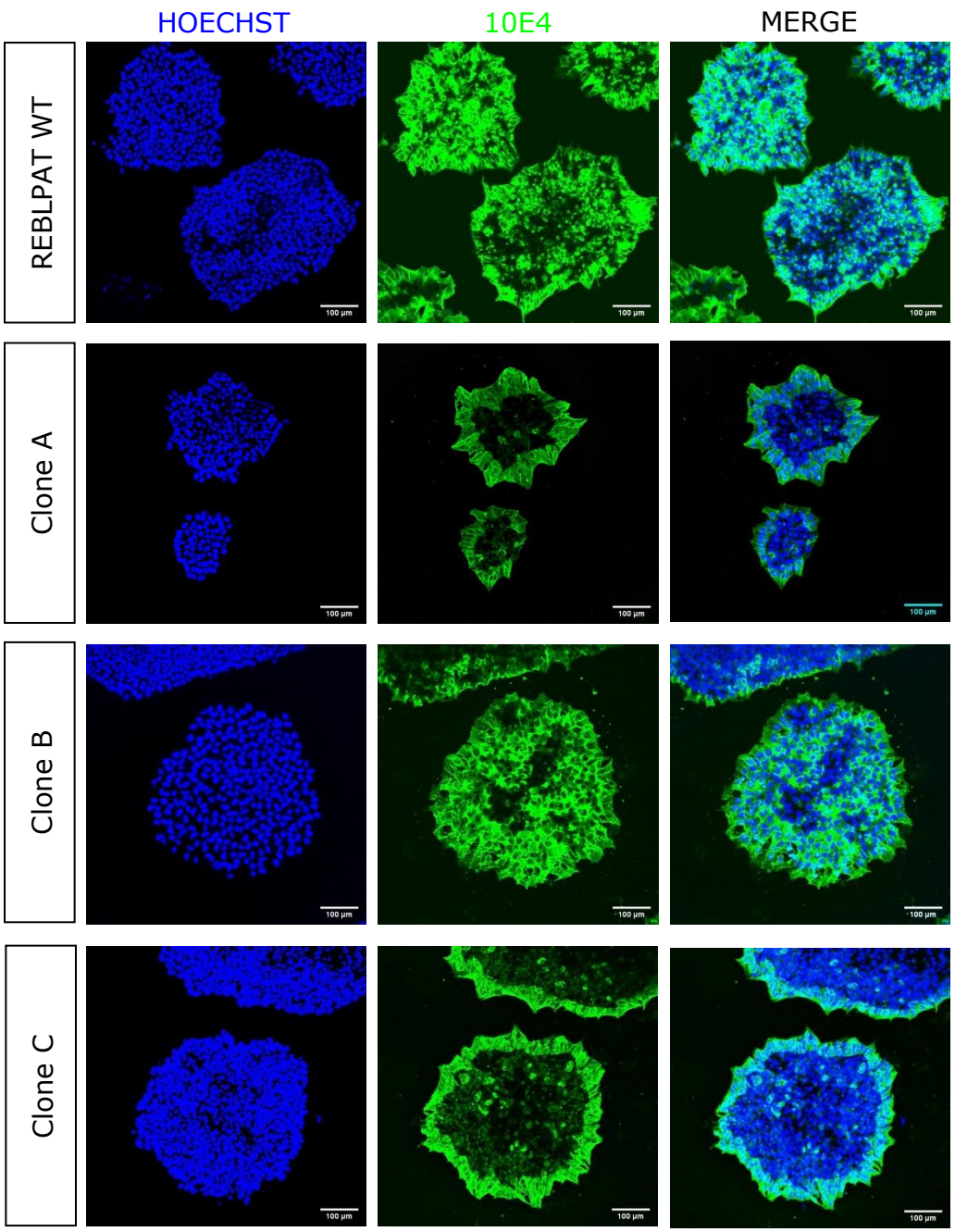


Figure 21 – Genetically Altered hiPSCs Produce Heparan Sulphate. Immunofluorescent images showing expression of 10E4 (HS) in three different clones (A, B, C) and REBLPAT WT control. 10E4 staining in all three clones appears similar to WT. Scale bars 100 μm

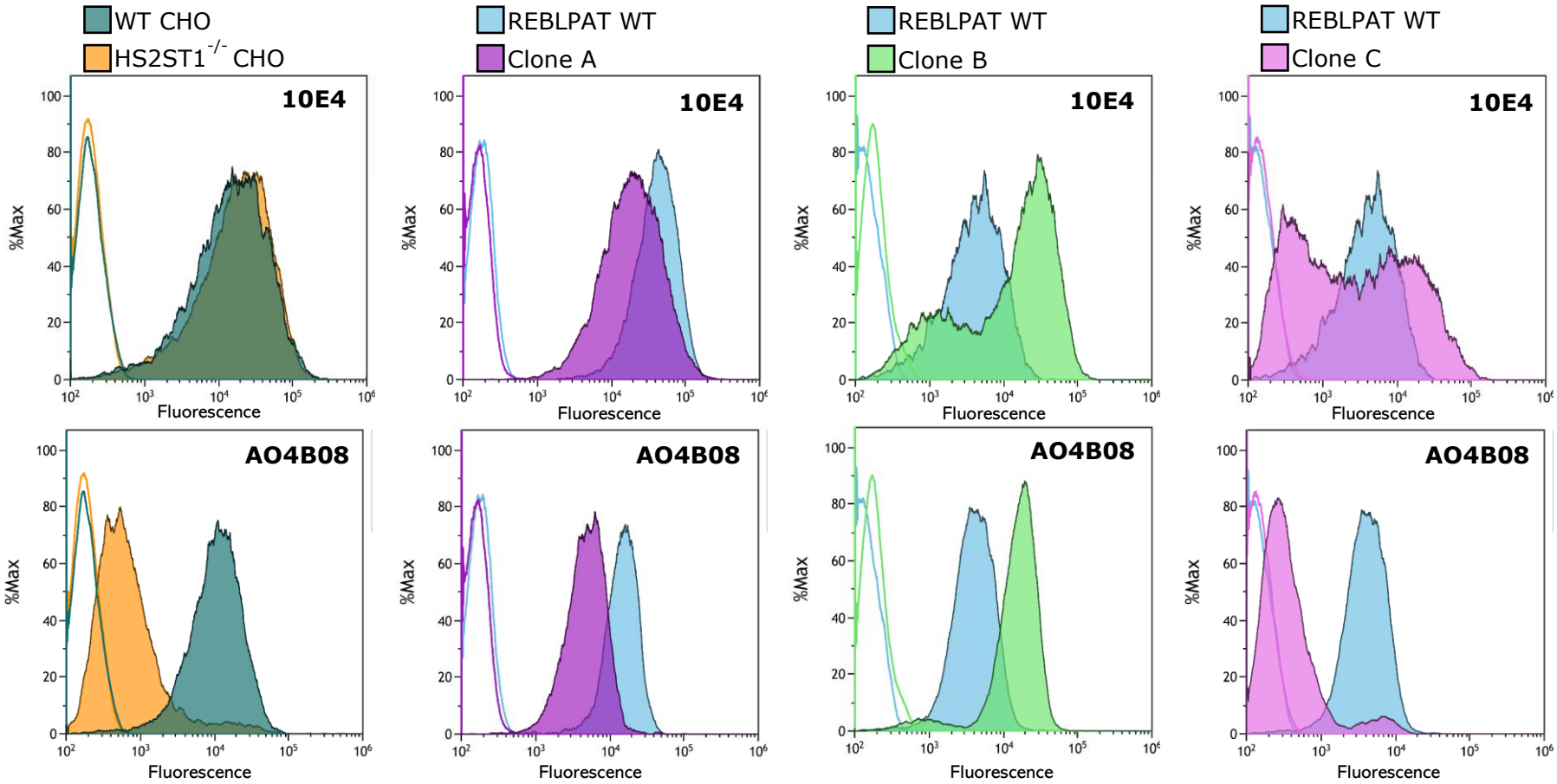


Figure 22 – Knockout Confirmation Using Flow Cytometry Cells were cultured in standard medium (CHO = Ham's/F12, hiPSCs = E8) for three days before carrying out flow cytometry. 10E4 stains heparan sulphate and plots show all cell lines still produce HS. AO4B08 stains 2-O-sulphated HS and indicates Clone A and Clone C have altered HS 2-O-sulphation. Outlined plots represent unstained controls. N =1 biological replicate. % Max was chosen for the y axis as this unit scales and normalizes each peak to 100%.

The AO4B08 antibody did not bind to the knockout CHO cell line (*HS2ST1*$^{-/-}$ CHO) but positive staining could be seen for the WT CHO cells, when the antibody was used at optimal concentration, as assessed by flow cytometry (**Figure 22**). hiPSC Clone C shows a clear difference in antibody binding when compared to REBLPAT WT control hiPSCs, similar to the difference seen between WT and KO CHO lines. Clone A also appears to have altered AO4B08 binding, but not to the same levels as observed in Clone C as determined by the shift in fluorescence peaks. Clone B shows high levels of AO4B08 staining with levels appearing to be marginally greater than those seen in the hiPSC control cell line (**Figure 22**). These findings were corroborated by the results of immunostaining. As expected, *HS2ST1*$^{-/-}$ CHO cells did not stain positive for AO4B08 (**Figure 23**). Similarly, Clone A and Clone C showed altered levels of AO4B08 staining and therefore HS 2-O-sulphation levels when compared with REBLPAT WT hiPSCs. Clone B showed similar levels of staining to WT (**Figure 24**).

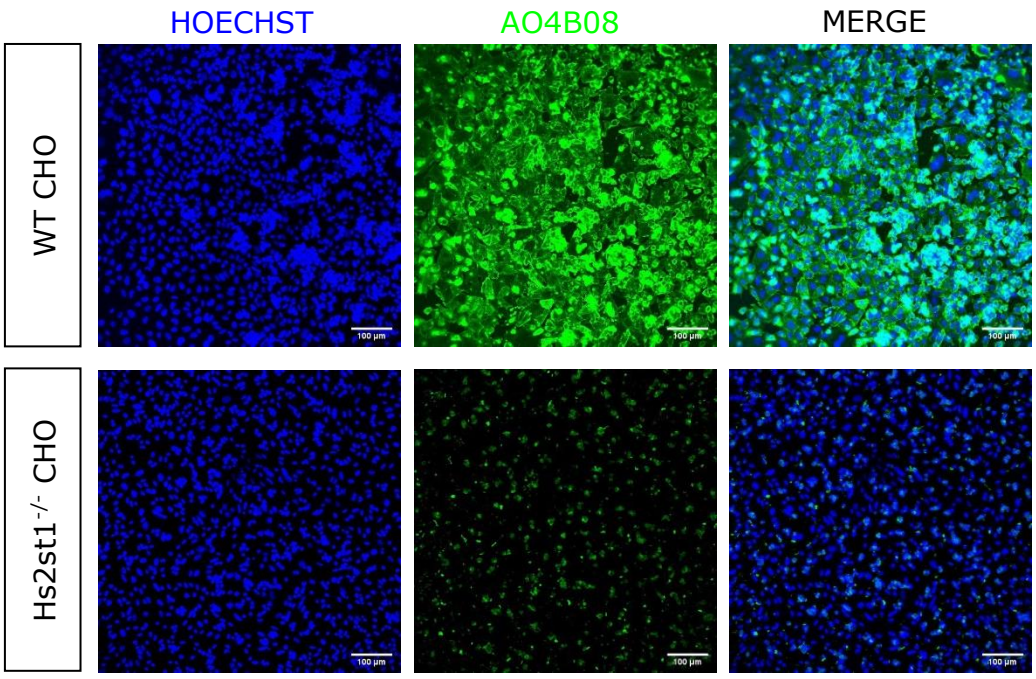


Figure 23 – Hs2st1$^{-/-}$ CHO Cells Do Not Stain for AO4B08. Immunofluorescent images showing expression of AO4B08 (2-O-sulphated HS) in WT and *Hs2st1*$^{-/-}$ CHO cell lines. AO4B08 staining is absent in *Hs2st1*$^{-/-}$ cells but present in WT. appears similar in both lines. Scale bars 100 μm

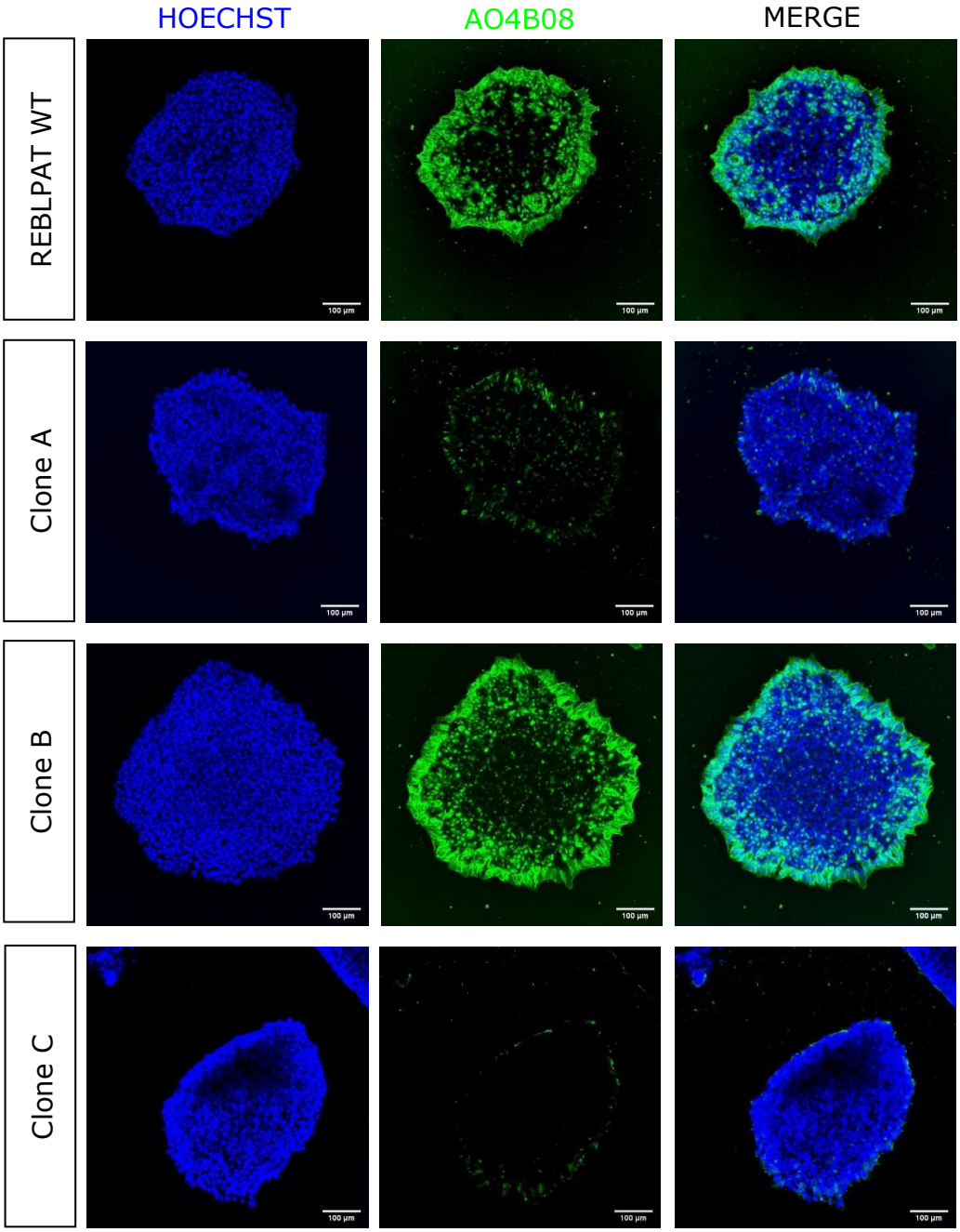


Figure 24 – Knockout Confirmation Using AO4B08 and Immunocytochemistry. Immunofluorescent images showing expression of AO4B08 (HS 2-O-sulphation) in three different clones (A, B, C) and REBLPAT WT control. Clone A and Clone C show staining different from WT while Clone B shows similar staining to WT hiPSCs. Scale bars 100 µm

4.3.5 hiPSC Lines with Reduced 2-O-Sulphation Have Increased 6-O-Sulphation

As expected, wild type controls (REBLPAT hiPSCs and CHO K1) display all of the detectable sulphation modifications (3-O-sulphation is unable to be detected by RP-HPLC due to a lack of available standards) while HS2ST1$^{-/-}$ CHO cells are lacking the 2-O-sulphated disaccharides. Similarly, Clone C also displays a total lack of 2S while Clone A has a very small contribution of 2S across the disaccharides – 3.2% and 2% in cell lysates and medium respectively. Clone A and Clone C also have an increase of 23% to 28% in the 6-O-sulphation modification compared to REBLPAT WT hiPSCs. Interestingly, when compared with WT CHO cells, *HS2ST1*$^{-/-}$ CHO cells did not show increased levels of 6-O-sulphation (**Figure 25 & 26**).

Focusing on the disaccharides in greater detail, the 2S modification in Clone A is confined to the NS residues with UA2S-NS(6S) levels below the detectable limit in both cell lysates and media. Furthermore, the abundance of the UA-GlcNS(6S) disaccharide is increased by 31 to 37% in Clone A and Clone C when compared with WT hiPSCs. Levels of UA-GlcNS(6S) were not observed to increase in HS2ST1$^{-/-}$ CHO HS but an increase of between 21% and 39% was detected in UA-GlcNS levels in HS from both cell lysates and conditioned medium (**Figure 25 & 26**). The overall modifications and HS disaccharide composition was similar between Clone B and REBLPAT hiPSCs (**Figure 25 & 26**).

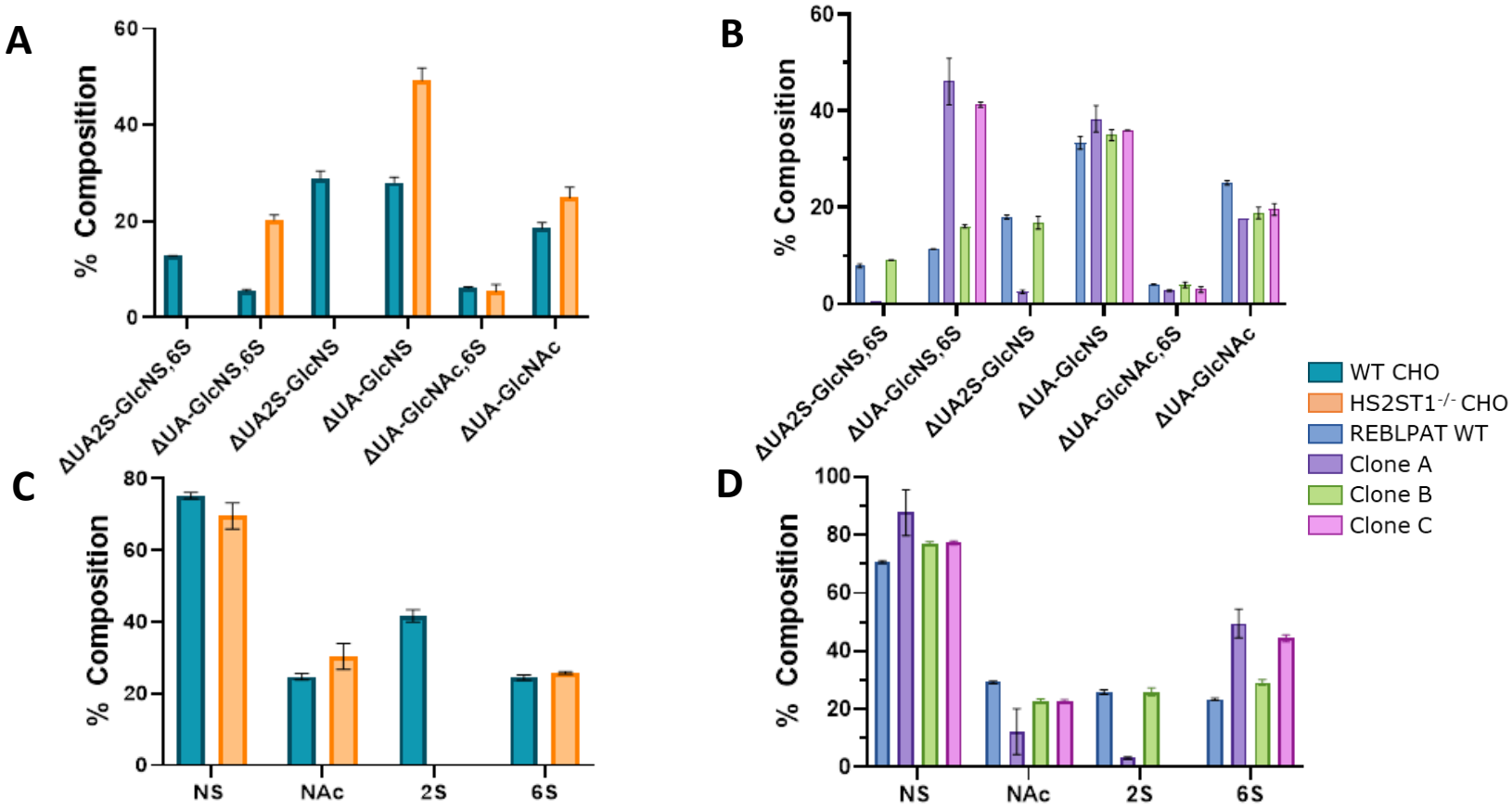


Figure 25 – Heparan Sulphate Disaccharide Composition of Cell Lysates. Cells were cultured under standard conditions (CHO = Hams/F12, hiPSCs = E8 and VTN-N) for three days. Cells were lysed and followed by HS disaccharide composition analysis using AMAC labelling and RP-HPLC. Note that UA corresponds to either GlcA or IdoA due to loss of stereochemistry after enzymatic degradation. A) Detailed HS disaccharide composition of CHO cells B) Overall HS modifications of CHO cells C) Detailed HS disaccharide composition of hiPSCs D) Overall HS modifications of hiPSCs. Error bars represent ± SD from n=2 technical replicates. N=1 individual experiment.

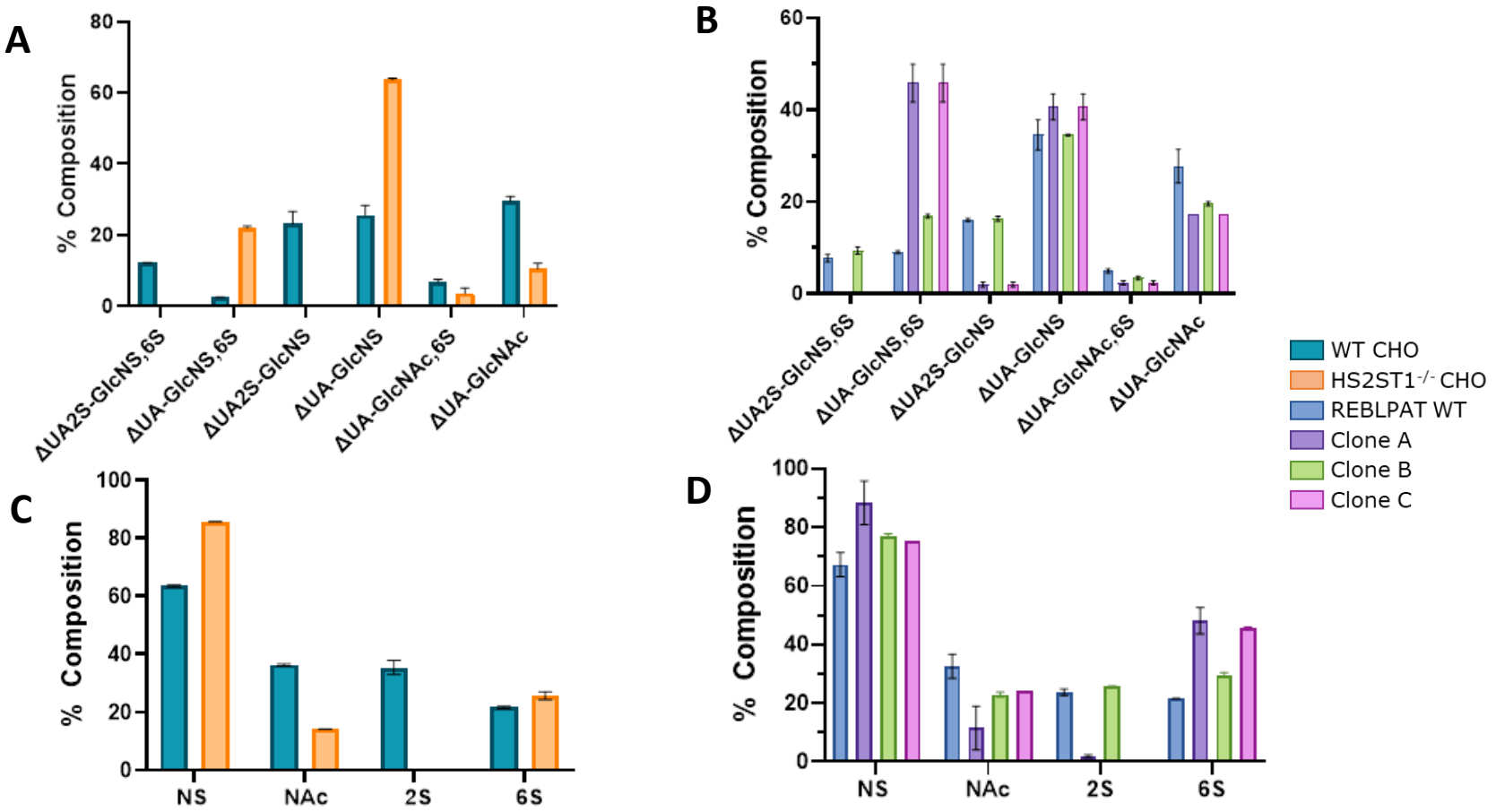


Figure 26 – Heparan Sulphate Disaccharide Composition of Condition Medium. Cells were cultured under standard conditions (CHO = Hams/F12, hiPSCs = E8 and VTN-N) for three days. Conditioned medium was collected daily and HS disaccharide composition analysis using AMAC labelling and RP-HPLC. Note that UA corresponds to either GlcA or IdoA due to loss of stereochemistry after enzymatic degradation. A) Detailed HS disaccharide composition of CHO medium B) Overall HS modifications of CHO medium C) Detailed HS disaccharide composition of hiPSC medium D) Overall HS modifications of hiPSC medium. Error bars represent ± SD from n=2 technical replicates. N=1 individual experiment.

4.4 Discussion

The work detailed in this chapter describes the HS compositional analysis of material obtained from patients who were the first ever to be identified with a loss of HS2ST1 function. In addition, the creation of an hiPSC-based model of the disease, generated by a CRISPR/Cas9 knockout of the *HS2ST1* gene is discussed. At the start of this study there were no reports of any human disease associated with loss of HS2ST1 function and, although animal models of loss of HS2ST1 function exist, no hiPSC models have been reported (Bai and Esko, 1996; Bullock *et al.*, 1998; Merry *et al.*, 2001). The discovery of patients who have a disease phenotype associated with *HS2ST1* mutations provided a clear need to generate a tool to assist in understanding the molecular basis of the novel disease.

4.4.1 A Comparison of Mouse Models and Human Patients

Mice with a homozygous gene trap insertion of *Hs2st1* are non-viable and die in the prenatal period or within 24 hours of birth. This is likely due to bilateral kidney agenesis, as all other internal organs appear normal. Bone mineralization was shown to be increased in homozygotes which also displayed polydactyly. Various eye defects were also seen in homozygous knockouts including colobomas, cataracts and stunted eye development (Bullock *et al.*, 1998).

In good agreement with the *Hs2st1*$^{-/-}$ mouse phenotype, all four patients identified by the Kutsche Lab at the Institute of Human Genetics in Hamburg were reported to have various skeletal abnormalities and three patients had issues with their vision. Where appropriate, the vision of patients was assessed, and of the patients, all three displayed abnormalities. Patient 1 had visual impairment while patient 2 was short sighted and patient 3 had various issues including underdeveloped irises and cataracts. The skeletal abnormalities varied between the patients but included craniofacial dysmorphisms, multiple flexion contractures and reduced ossification. All patients had issues with their hands and feet, namely broad phalanges, in-turned feet or flat feet and brachydactyly or camptodactyly. One out of the four patients displayed bilateral renal agenesis (as occurs in the mice) suggesting that the impact of a loss of HS2ST1 activity leads to similar developmental defects across species. However, patients 2

and 4 were found to have unilateral renal agenesis while patient 1 had no renal abnormalities (Schneeberger *et al.*, 2020). The differences in kidney development could indicate that these patients have varying levels of HS2ST1 enzyme activity and therefore varying levels of HS 2-O-sulphation resulting in different phenotypes. As mice with homozygous gene trap mutations in *Hs2st1* have been shown to have no detectable HS 2-O-sulphation and bilateral renal agenesis, it could be hypothesized (prior to HS disaccharide compositional analysis) that the mutation in patient 1 did not lead to total loss of function, and therefore renal development could still occur.

In order to ascertain whether *HS2ST1* mutations in patients led to loss of function, the HS disaccharide composition was determined. Only HS from patient 1 was able to be analysed as the amount of HS extracted from the other patients' cells was, unfortunately, insufficient to generate a valid HPLC read-out. Although low levels of 2S were detected in the HS from patient 1, it should be noted that the cells were cultured in serum containing medium (essential for primary fibroblast culture) which showed detectable levels of 2S so the observed 2S is likely to be an artefact, rather than reflecting HS2ST1 activity. The overall sulphation patterning of HS in patient 1 is similar when compared with the HS from *Hs2st1*$^{-/-}$ mice with increased 6S and NS. Looking in detail, there was a significant increase in UA-GlcNS(6S), while levels of UA-GlcNAc(6S) remained the same in both the patients and mice, showing that the increase in 6S is confined to the NS residues. Although HS could not be purified from the other patients, the analysis of the HS from patient 1 suggests that the range of phenotypes seen in the patients is not due to varying degrees of enzyme activity. Although it is clear that there is only one HS 2-O-sulphotransferase enzyme (in all vertebrates studied to date), the data presented here confirms that mutation of the HS2ST1 enzyme in humans results in total loss of 2S within HS. This provides additional evidence that *HS2ST1* loss is not compensated for by any other 2-O-sulphotransferase family members. The reason, and mechanism, behind the increase in NS and 6S could be a result of various things. The loss of HS2ST1 adding sulphate group onto the HS chain creates and abundance of free sulphate group within the Golgi that could be added onto the HS chain. This could result in the NDST and HS6ST enzymes increasing their activity as a result of the increase in available

substrate. Another, perhaps less plausible, possibility is that due to loss of the HS2ST1 enzyme, the amount of NDST and HS6ST enzyme present are increased. However, this would require genetic regulation and increased transcription and translation to produce a greater amount of the enzyme itself. The hierarchical order of HS biosynthesis in **Figure 4** is clearly subverted when an essential enzyme is missing. Though this is seen across several species, there is no known reason, or hypothesis, for the occurrence. Despite the clear subversion of the hierarchy, it is still widely believed that this exists and the abnormal HS seen in mutants is as a result of the change in HS biosynthetic machinery. The results shown presently, and in the mouse model, indicate that the HS6ST enzymes do not require the HS2ST1 enzyme to act on the Hs chain first. Therefore, it could be said that there isn't a strict hierarchical order of HS biosynthesis with reference to the sulphotransferases.

Analysis of the HS from patients 2, 3 and 4 is a high priority to allow conclusions to be drawn regarding genotype/phenotype correlation for the novel disease. Currently, it is hypothesized that the HS will appear similar to that of patient 1 and the mutant mice and that there will be an overall loss of 2S, coupled with an increase in NS and 6S. Furthermore, culturing the cells in serum free medium would ensure that there is no contamination with exogenous HS added and will allow for a clear conclusion regarding the presence of the 2S. Ideally, HS disaccharide compositional analysis of patient 1 would also be repeated using cells cultured in serum-free medium. Additionally, other methods of analysis can be used to investigate the effects of the KO on HS composition including ToF-SIMS, introduced in Chapter 1.

4.4.2 Implications of Loss of HS2ST1 on Humans

HS2ST1 is widely expressed from early embryogenesis and 2-O-sulphation appears to be a ubiquitous modification within HS, with all native HS analysed to date containing at least some 2S (Maccarana *et al.*, 1996; Bullock *et al.*, 1998; Lin *et al.*, 2000). However, it has been established that a complete loss of HS results in failure of gastrulation (Lin *et al.*, 2000; Stickens *et al.*, 2005). Therefore, it is somewhat surprising that mice lacking 2S within their HS have normal development of most organs, with only the kidneys displaying a completely penetrant phenotype. It is therefore even more surprising that over 20 years after the

generation of the mouse models, humans have been discovered living without the HS2ST1 enzyme.

In the four patients discovered to have alterations in the *HS2ST1* gene, only three had kidney abnormalities. Patient 1 has been shown to have significantly reduced HS 2-O-sulphation and therefore it could be assumed reduced FGF2 binding sites. However, near normal development, particularly of the kidneys, has still occurred. Patients 3 and 4 have identical homozygous mutations and so it can be hypothesized they have identical levels of HS2ST1 and similar HS compositions. However, patient 3 displayed bilateral renal agenesis while patient 4 only has unilateral renal agenesis. It is possible that to elicit a downstream response, FGF2 does not always need high affinity binding and therefore 2-O-sulphation may not always be necessary and is instead tissue or cell specific. It has previously been shown that GAGs isolated from giant African snails, *Achatina fulica,* have a high binding affinity for FGF2, but low mitogenic activity when compared to a heparin control. It was thought that the mitogenic effects seen in the control were as a result of 6-O-sulphate groups not present in the GAGs isolated from the snail, indicating that the binding affinity of FGF2 to HS does not have an effect on the downstream response (Wang *et al.*, 1997). Furthermore, the theory that maintained charge density (due to increased NS and 6S) allows near normal cellular function and embryonic development is possible as this is something seen across both species. This, however, does not account for the variation among renal phenotypes. It is possible that Patient 1 still has some residual HS2ST1 activity, but confirmation of this is not possible without culturing cells in serum-free and GAG-free medium. If increased NS and 6S motifs allow for FGF2 binding and ERK phosphorylation, it could be possible that the increase of these motifs in patient 1 is higher than that seen in mouse models and the other patients which has allowed for ureteric bud bifurcation and metanephric mesenchyme condensation. This would provide an explanation of the differences seen between patient 3 and patient 4. Patient 4 may have higher levels of NS and 6S which has allowed for the development of a single kidney, while patient 3 has a minimal increase resulting in bilateral renal agenesis. However, to confirm this, HS disaccharide compositional analysis would need to be performed (preferably after culture in serum-free media).

4.4.3 Are hiPSCs That Do Not Display 2-O-Sulphation Knockouts?

As previously explained, antibodies able to detect endogenous levels of HS2ST1 enzyme do not exist, so the downstream products of the enzyme activity were detected to identify knockouts. In order to be confident about whether hiPSCs were HS2ST1 deficient, several methods were used. *Hs2st1*$^{-/-}$ and WT CHO cells were used as controls, along with WT hiPSCs. *Hs2st1*$^{-/-}$ CHO cells have previously been shown to stain negative for AO4B08 and so served as a negative control (Kurup *et al.*, 2007). To ensure that the absence of 2-O-sulphation (and negative AO4B08 staining) is not due to a total loss of HS, 10E4 was used to visualize HS. Due to the clear difference in binding of AO4B08, determined by both flow cytometry and immunostaining, Clone C was determined to completely lack detectable 2-O-sulphated HS. Clone A is likely to still have 2-O-sulphated HS. However, antibody binding appears altered when compared to wild type, suggesting that this clone could have a heterozygous genotype. This is confirmed by the compositional disaccharide analysis which found Clone C to have a total lack of 2-O-sulphation and Clone A to have reduced levels but some detectable 2S. Additionally, both clones show a similar sulphation pattern to the knockout mice and patient 1 – with a specific elevation in the UA-GlcNS(6S) disaccharide indicating that an increase in NS and 6S is key to maintaining near normal cell functions. Although CHO cells were used as a control, the same change in sulphation patterning was not observed here, with only an increase in the UA-GlcNS disaccharide detected and not the disulphated disaccharide. This was as expected because it has previously been shown that 6-O-sulphation levels are only slightly increased (by 4%) in *HS2ST1*$^{-/-}$ CHO cells when compared with wild type cells. Furthermore, the initial study in which mutant CHO cell lines were generated reported that N-sulphation increased by 19% in the mutant cell lines (Bai and Esko, 1996). Despite having genetic alterations, Clone B still appeared to have similar levels of HS 2-O-sulphation, and HS disaccharide composition, when compared to wild type hiPSCs, so it can be assumed that the genetic alterations have had no effect on the enzyme activity.

Unlike with the plasmid based, HDR CRISPR delivery strategy, the gene is not disrupted with an antibiotic resistance cassette. Therefore, it was impossible to tell whether clones were enzyme knockouts from sequencing alone. Furthermore, Clone A and Clone C show similar levels of *HS2ST1* mRNA which shows the hiPSC lines are not genetic knockouts as the gene transcript is not disrupted. Having said that, *Hs2st1* CHO cells are also not true genetic mutants as they are mutagenized with ethylmethane sulfonate but are often still classed as knockouts (Bai and Esko, 1996). Mouse models, despite being generated by gene trap methodologies, have disrupted *Hs2st1* transcripts (Bullock *et al.*, 1998). The homozygous mutation in Clone C results in a stop codon in Exon 2 so the absence of 2-O-sulphation in the HS is likely due to early arrest of protein translation, resulting in a truncated form of HS2ST1 that lacks the sulphotransferase domain. However, it has been reported that premature stop codons (or whole exons) can be skipped to produce partial functional gene product in which the protein is potentially truncated and slightly misfolded (Lalonde *et al.*, 2017; Mou *et al.*, 2017; Stauffer, Biswas and Sharan, 2020). This could potentially be the case with Clone C. However, from a functional perspective, it can still be said that the enzyme has no functionality due to loss of the ultimate enzyme product – 2-O-sulphation within HS. With Clone A, it is difficult to determine whether abnormal enzyme function is due to halted protein translation, or abnormal protein folding. As protein structure confers its function, correct folding will be essential for correct transfer of sulphates to the HS chain. If the protein is translated, but non-functional due to abnormal folding, anti-HS2ST1 antibodies would still not stain for the endogenous protein as the usual binding epitopes would likely be altered. Furthermore, traditional enzymatic assays that utilize colorimetric or radiometric readouts could also be used to determine HS2ST1 function (Paul *et al.*, 2012; Sterner *et al.*, 2014). However, these assays would not provide insight into whether the clones have abnormal 2-O-sulphation due to improper translation or protein folding. A more detailed investigation into the structure of the endogenous protein of each of the hiPSC lines would be required to determine the cause of loss of HS2ST1 function.

It is important to note that multiple knockout confirmation methodologies were chosen to mitigate the chances of false

antibody binding and as a result, incorrect identification of knockouts. As previously mentioned, AO4B08 preferentially binds to 2-O-sulphated regions of HS, but not exclusively. This is why a negative CHO control was used to titrate the antibody and determine optimal conditions. However, there was still a chance that in some cases AO4B08 could bind preferentially and provide a positive (or WT) read out; the reverse is also true. Flow cytometry and immunostaining were carried out at separate times on biological replicates of the cells, so although each was only done once, AO4B08 binding to the cells was carried out twice and a different detection method was used. Although compositional disaccharide analysis provided information regarding changes in the global sulphation patterning of the HS, it was still useful to determine whether hiPSC lines produced 2-O-sulphated disaccharides. Furthermore, the analysis was carried out on another passage of the hiPSCs, and thus a further biological replicate. Each of the knockout detection methods was only carried out once but as three were used in total, it provides similar data support as running one method three times. Therefore, the methods chosen here can be likened to three biological replicates.

4.4.4 Summary of CRISPR/Cas9 Targeting

It can be concluded from the findings discussed above that Clone B has no changes in 2-O-sulphation levels of HS and so likely has genetic alterations that still result in functional enzyme. As a result, this clone will not be included in any further investigations. As the *HS2ST1* gene is still intact, Clone A and C should not be referred to as genetic knockouts. Furthermore, the reason for loss of function can also not be determined meaning they should also not be referred to as protein knockouts. As only the downstream products of the enzyme have been investigated, any clones with abnormal HS 2-O-sulphation can only be referred to as functional knockouts. Altered HS 2-O-sulphation staining indicates that Clone A is a heterozygous functional knockout and Clone C is a homozygous functional knockout due to a complete absence of the 2S motif. Therefore, henceforth these clones will be referred to as HS2ST1$^{+/-}$ and HS2ST1$^{-/-}$, respectively. However, it is important to note they are only functional knockouts and no information regarding the presence of the protein has been determined. Where appropriate, it

will be indicated if HS2ST1$^{+/-}$ or HS2ST1$^{-/-}$ is referring to the loss of gene or protein rather than functionality. A summary of the conclusions drawn regarding the clones characterized in this chapter can be seen in **Figure 27**.

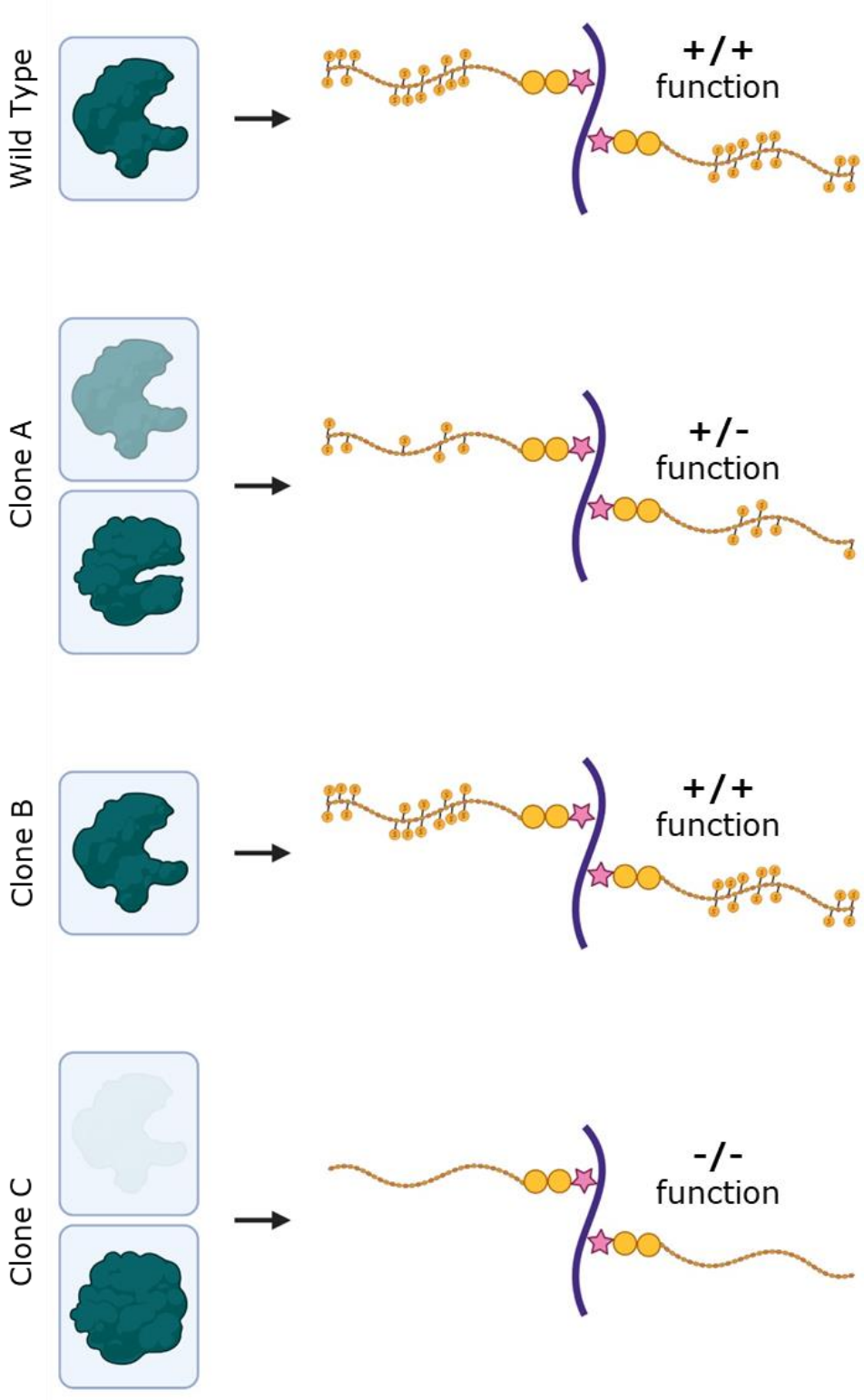


Figure 27 - HS2ST1 CRISPR/Cas9 Summary. Clone A has HS that displays reduced 2-O-sulphation which could be the result of reduced levels of translated enzyme or abnormal folding resulting in reduced activity. Clone B has wild type HS. Clone C has no 2-O-sulphation motifs on the HS due to either abnormal folding or no translation. It can be concluded that Clone A has heterozygous loss of HS2ST1 function and Clone C has homozygous loss of HS2ST1 function.

Chapter 5

Characterization of hiPSCs with Abnormal Heparan Sulphate

5.1 Introduction

5.1.1 Chapter Overview

The work described in this chapter aimed to investigate the differentiation capability of hiPSCs with altered heparan sulphate 2-O-sulphotransferase (HS2ST1) function (as described in Chapter 4). Initially, the pluripotency of the hiPSCs was assessed before directed differentiation to endoderm and ectoderm (neural) lineages was carried out. Heparan sulphate is known to interact with several signalling factors such as FGF2, Wnt and TGFβ to mediate pluripotency and differentiation (**Figure 10 and 11**), so this chapter will investigate whether loss of HS2ST1 activity affects these key cellular processes.

As the main aim was to assess whether HS2ST1$^{-/-}$ hiPSCs would differentiate to all three germ layers; endoderm, ectoderm and mesoderm, protocols for these had to be selected. The endoderm differentiation protocol chosen is routinely used at the University of Nottingham and so was prioritised based on the availability of local scientific expertise. Furthermore, hiPSCs lacking HS have previously been differentiated to definitive endoderm so the differentiation of these hiPSCs with altered HS can be directly compared. To investigate ectoderm, a neural differentiation was selected due to the neurological impairments seen in the patients. At the time of this study, no GAG-free neural differentiation was routinely used at the University of Nottingham. The differentiation protocol was chosen based on literature and product availability. However, further optimization of the neural differentiation is required to be able to fully characterize the effects of the functional knockout on neural development. Mesodermal differentiation could not be completed in the timeframe of this project.

It was originally hypothesised that pluripotency and differentiation would be affected due to loss of HS2ST1. However, since work on

HS deficient cell lines has been carried out (section 1.3.2), we now hypothesise that pluripotency and differentiation to endoderm will not be affected. The original hypothesis of the *EXT1* and *EXT2* work in hiPSCs was that the cells would be stuck in a pluripotent state, as described in mESCs (Johnson *et al.*, 2007; Kraushaar, Yamaguchi and Wang, 2010; Holley, Pickford, *et al.*, 2011; Pickford *et al.*, 2011). However, FGF2 is required to maintain pluripotency in hiPSCs so it could be thought this would also have beeb altered (Apáti, Berecz and Sarkadi, 2016). A summary of the differentiation protocols used in this doctoral thesis can be seen in **Figure 28**.

5.1.2 Definitive Endoderm Differentiation

Definitive endoderm (DE) gives rise to epithelial cell types that are found lining most of the organs in the respiratory and digestive systems, while its patterning also gives rise to organs such as the thyroid, liver and pancreas (Wells and Melton, 1999). During early gastrulation, the primitive streak forms in the epiblast and, as cells migrate through the primitive streak, they form either endoderm or mesoderm. During *in vivo* embryonic development, high levels of Nodal signalling are required for DE formation, and DE forms close to the Nodal source, which in mice is the anterior primitive streak (Zorn and Wells, 2009). Wnt ligands are a large family of secreted glycoproteins that interact with specific receptors to activate the canonical Wnt/β-catenin and non-canonical β-catenin independent signalling pathways (Inestrosa and Varela-Nallar, 2015). Wnt3, a member of the Wnt family of ligands, is expressed during early gastrulation, and studies have shown that Nodal and Wnt3 mutually activate each other, and therefore both are required for DE formation (Grapin-Botton and Constam, 2007). DE is characterized by the presence of transcription factors such as SOX17, FOXA2 and GATA4. These transcription factors, whose presence is promoted by Nodal signalling, cause a cascade of downstream effects that commit cells to the endodermal lineages (Zorn and Wells, 2009). *In vitro*, DE can be formed from hPSCs in four days by using medium supplemented with Activin A and Wnt3a (Jacob *et al.*, 2019). Differentiation can then be taken further into more specific endodermal lineages. As *in vitro* DE formation requires Activin A and Wnt3a, alterations in HS are likely to affect its formation due to the requirement for HS to sequester ligands

involved in these signalling pathways. A more in-depth discussion into HS and signalling pathways can be found in Chapter 1.

5.1.3 Ectoderm (Neural) Differentiation

The ectodermal germ layer gives rise to the entire nervous system and various glandular tissues (Sasai and De Robertis, 1997). It was originally thought that the commitment of ectoderm cells towards the neural fate relies on signals from other adjacent tissues. Early work on *Xenopus* embryos indicated that a cluster of dorsal mesodermal cells, known as the organiser, release signals that induce neural differentiation in the ectoderm (Ruiz i Altaba, 1994; Spemann and Mangold, 2001). A gradient of Wnt is essential for neural plate axis specification in the early embryo. Not only is Wnt signalling in the posterior necessary, but also its inhibition in the anterior by Dickkopf1 (Dkk1) is required for correct patterning (Kiecker and Niehrs, 2001). Later in development, Wnt signalling is required for differentiation of neural precursor cells (NPCs) to more mature neural cells via inhibition of their self-renewing capacity (Hirabayashi *et al.*, 2004). Dysregulation of Wnt signalling can lead to developmental issues and brain malformations including cortical dysplasia (Munji *et al.*, 2011). The formation of NPCs in hiPSCs with functionally altered HS2ST1 may be different from wild type due to the role of HS in Wnt ligand dispersal (discussed in section 1.7.2).

As well as signals that actively induce neural formation, other research indicates that the organiser also releases inhibitory factors. These inhibitors result in the absence of signals that would ordinarily induce the cells to form other lineages, but the absence of such signalling allows commitment of cells to the neural fate (Hemmati-Brivanlou and Meltont, 1997). These inhibitory factors include Noggin, Chordin and Follistatin which bind a group of growth factors known as bone morphogenetic proteins (BMPs) to inhibit epidermal differentiation (Lamb *et al.*, 1993; Sasai *et al.*, 1995; Mulligan and Cheyette, 2012). The tightly regulated activation and inhibition of BMPs helps to drive neural development. However, these BMP antagonists are not enough to induce neural formation alone, fibroblast growth factor (FGF) signalling is also a requirement (Stern, 2005). Smad proteins are part of the BMP signalling pathway and their activity and

translocation to the nucleus is inhibited by phosphorylation. Mitogen-activated protein kinase (MAPK) phosphorylates Smad in the presence of FGF, thereby inhibiting BMP signalling. FGF negatively regulates Smad phosphorylation which provides an explanation as to why FGF and BMP appear to have opposing effects (Pera *et al.*, 2003; Liu and Niswander, 2005; Bond, Bhalala and Kessler, 2012).

Several *in vitro* neural differentiation protocols exist, many of which work by manipulating and recreating signalling pathways. Dorsomorphin or Noggin and SB431542 (a small molecule inhibitor of TGFβ/Smad signalling) are regularly used to inhibit BMP and Activin A/Nodal pathways to drive NPC commitment and prevent endoderm/mesoderm formation (Gerrard, Rodgers and Cui, 2005; Kim *et al.*, 2010). After NPC formation, neurons are often expanded in supplemented neurobasal medium to aid with the maturation of NPCs to neurons (Li *et al.*, 2020).

Neural differentiation protocols frequently require the use of Matrigel®, or laminin derived from Matrigel®. These coating substrates are unsuitable for the present study as they contain exogenous GAGs and could therefore counteract the effect of the HS2ST1 knockout that has been generated. For this reason, a commercially available neural supplement was chosen for the differentiation and the protocol could be easily supported by culture of recombinant vitronectin, which has been confirmed as GAG free (Jamie Thompson, personal communication). For expansion of NPCs to neurons, a protocol published by ThermoFisher was adapted to use a recombinant form of laminin in order to be GAG-free (Thermo Fisher Scientific Inc., 2015). Laminin-111 is commonly found in neural tissues and is essential for neurodevelopment (George and Geller, 2018; Halder, Sapkota and Milner, 2022). As recombinant laminin-111 is commercially available, it was hypothesized that this would be an effective substrate for neural maturation and was therefore used.

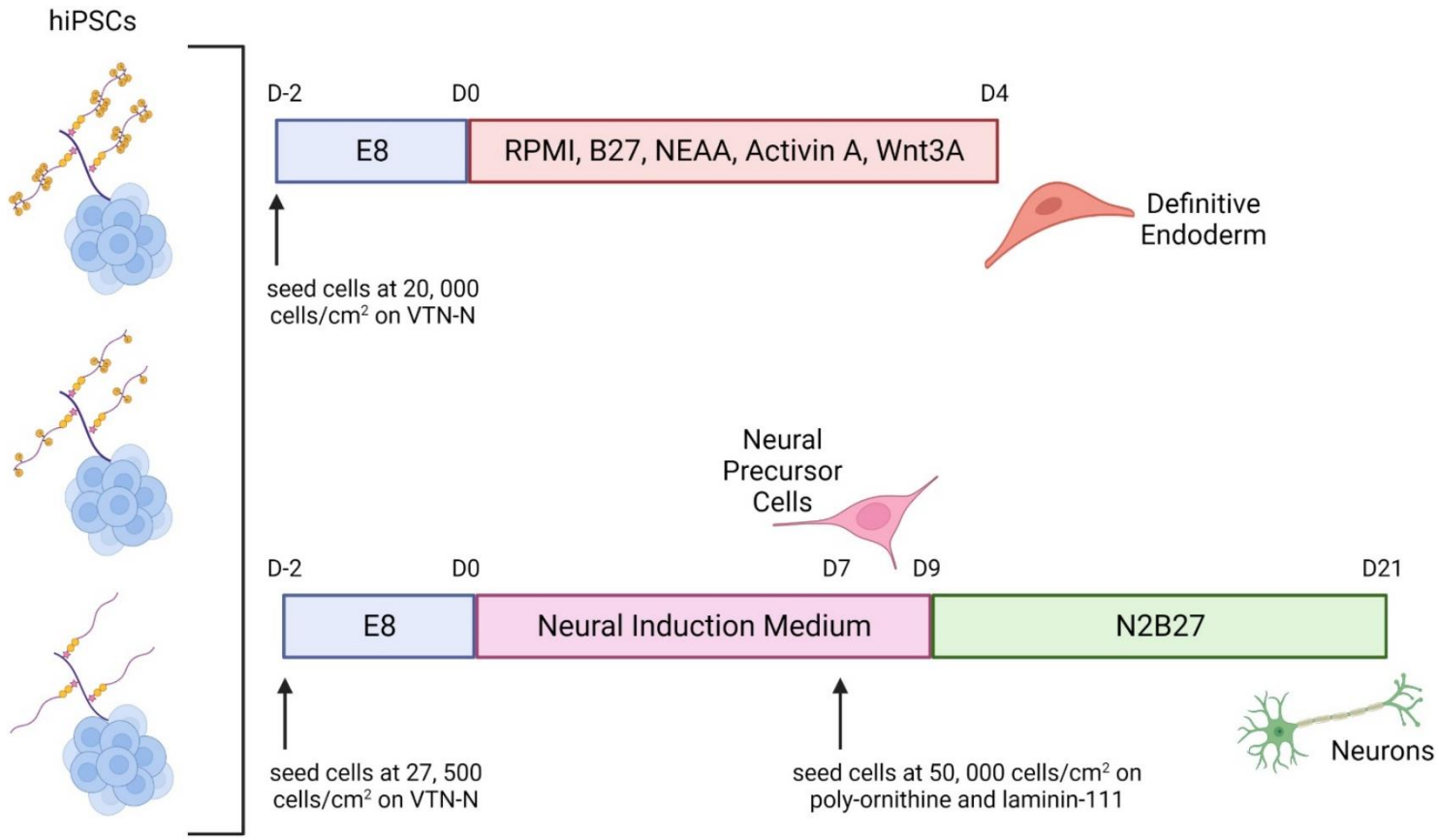


Figure 28 - Summary of Differentiation Protocols. hiPSCs were seeded at 20,000 cells/cm^2 on VTN-N and differentiated in RPMI supplemented with Activin A and Wnt3a for four days to generate definitive endoderm cells. hiPSCs were seeded at 27,500 cells/cm^2 on VTN-N and cultured in neural induction medium (unknown composition) for 7 days to generate neural precursor cells (NPCs). NPCs were then seeded onto laminin-111 and poly-L-ornithine at 50,000 cells/cm^2 and cultured in N2B27 for 14 further days to generate neurons.

5.2 Materials and Methods

5.2.1 Definitive Endoderm Differentiation

hiPSCs were seeded onto VTN-N at 20,000 cells/cm^2 in E8 medium supplemented with 10 µM ROCKi. Medium was changed to fresh E8 after 24 hours. 48 hours after seeding, medium was changed to DE differentiation medium: RPMI 1640 (Invitrogen, 251875-034) supplemented with 2X B-27™ (Invitrogen, 17504-044), 1X MEM Non-Essential Amino Acids Solution (NEAA; Gibco, 11140050), 100 ng/µL Activin A (R&D Systems, 338-AC-050) and 50 ng/mL Wnt3A (R&D Systems, 5036-WN-010). Medium was changed daily for four days.

5.2.2 Definitive Endoderm Flow Cytometry

Flow cytometry was carried out daily from day 1 to day 4 of DE differentiation. DE medium was aspirated, and the cells washed with DMEM/F12. Cells were then dissociated with Gentle Cell Dissociation Reagent (GCDR; Stem Cell Technologies, 100-0485) at 37°C for 5 minutes, collected and pelleted at 300 x g for 5 minutes. Cells were resuspended in 300 µL DE FACS buffer (2% FBS in 1X DPBS) and split three ways (100 µL per condition – unstained, isotype control (co-stained) and DE markers (co-stained)). Antibodies were diluted in DE FACS buffer and incubated with cells for 30 minutes on ice, in the dark. Antibody details can be found in **Table 14**. Cells were washed in DE FACS buffer twice, pelleting between each wash. The cell pellets were resuspended in 500 µL DE FACS buffer and transferred to an ID7000™ Spectral Cell Analyzer (Sony) for analysis. Flow pressure was set at 25% and a total number of 20,000 events were recorded (in the single cells gate) for each sample. Resulting data was analysed using Kaluza software (Beckman Coulter).

Table 14 - Antibodies Used for Definitive Endoderm Flow Cytometry. CXCR4 encodes a chemokine receptor known as CD184 while cKit encodes a tyrosine protein kinase known as CD117.

Antibody	Isotype	Dilution	Supplier Details
CXCR4-PE	Mouse, IgG2a	1:50	Life Technologies, MHCXCR404
cKit-APC	Mouse, IgG1	1:20	Invitrogen, 17-1179-42
Isotype Control-PE	Mouse, IgG2a	1:50	Invitrogen, 12-4724-82
Isotype Control-APC	Mouse, IgG1	1:50	Invitrogen, MG105

5.2.3 Neural Precursor Cell Differentiation

hiPSCs were seeded at 27,500 cells/cm^2 onto pre-coated VTN-N plates in E8 medium supplemented with 10 µM ROCKi. After 48 hours, medium was changed to PSC Neural Induction Media (Gibco, A1647801). Medium was changed every other day for seven days, until NPCs had formed. The morphology of the cells changed to densely packed cells with some 3D areas giving the impression of an undulating effect in the plates.

5.2.4 Neural Precursor Cell to Neuron Differentiation

Glass coverslips were coated with poly-L-ornithine (Sigma, P3655) diluted to 20 µg/mL in sterile water overnight at 4°C. The coverslips were washed twice with 1X DPBS and coated with laminin-111 (Biolamina, LN111) diluted in 1X DPBS with Ca^{2+}/Mg^{2+} (Gibco, 14040141) to a final concentration of 1 µg/cm^2 for 2 hours at 37°C. NPCs were dissociated using accutase solution for 5 minutes at 37°C and collected in NI medium. Cells were pelleted at 300 x g for 5 minutes and resuspended in fresh neural induction medium and a single cell suspension generated. NPCs were seeded at 50,000 cells/cm^2 onto pre-coated glass coverslips in NI medium supplemented with 10 µM ROCKi. After 48 hours the medium was changed to N2B27 medium: 1:1 ratio of Neurobasal medium (Life Technologies, 25030-024) and DMEM/F12 (Sigma, D8062) supplemented with 1X N2 supplement (Life Technologies, 17502-

048), 1X B27 supplement (Life Technologies, 17504-044) and 1X L-glutamine (Life Technologies, 25030-024). Medium was changed every 2 days for a further 14 days.

5.3 Results

5.3.1 hiPSCs With Functionally Altered HS2ST1 Express Pluripotency Markers

As HS 2-O-sulphation is necessary for FGF2 signalling, and FGF2 is required to maintain pluripotency, it was hypothesized that hiPSCs with altered HS2ST1 function would display differences in pluripotency when compared with wild type hiPSCs (Turnbull *et al.*, 1992; Apáti, Berecz and Sarkadi, 2016). Pluripotency was assessed by qRT-PCR and immunostaining of NANOG and OCT4, transcription factors known to play a critical role in maintaining hiPSCs in an undifferentiated state. KO cell lines were compared against wild type control hiPSCs (REBLPAT WT) known to have no alterations in HS. Surprisingly both $HS2ST1^{+/-}$ and $HS2ST1^{-/-}$ expressed NANOG and OCT4 at relatively similar intensity levels to REBLPAT WT cells when assessed by immunofluorescence, although this has not been quantified (**Figure 29A**). Similar results were observed via qRT-PCR; both hiPSC lines displayed similar levels of NANOG and OCT4 gene expression when compared with WT controls (**Figure 29B**).

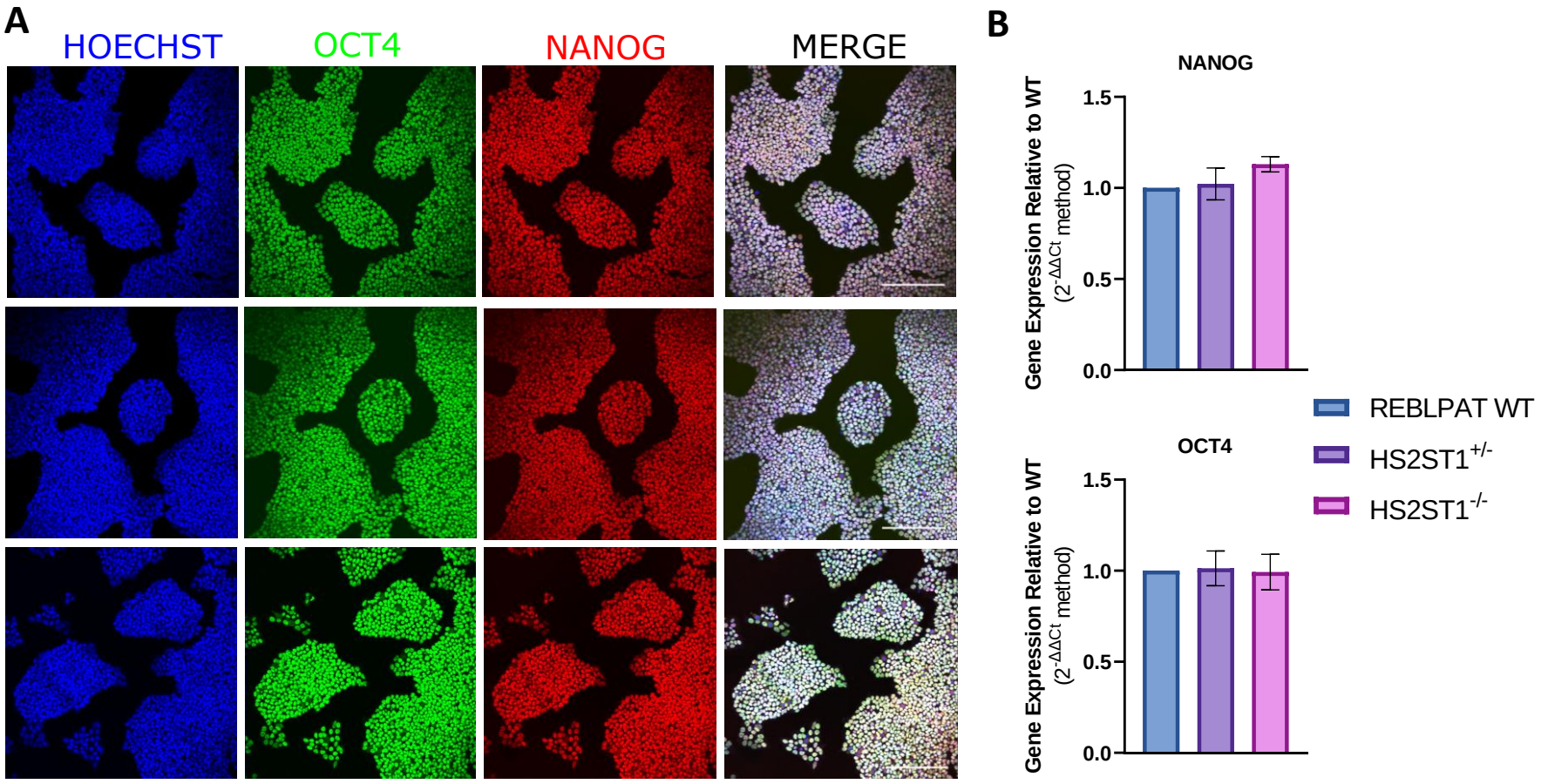


Figure 29 - HS2ST1$^{+/-}$ and HS2ST1$^{-/-}$ hiPSCs Express Pluripotency Markers. A) Immunofluorescence images and B) gene expression analysis of NANOG and OCT4 pluripotency markers. Gene expression was normalized to PBGD. hiPSCs were cultured on VTN-N in E8 medium for 3 days. Error bars represent ± SEM from N=3 independent experiments; scale bars 200 µm

5.3.2 hiPSCs with Functionally Altered HS2ST1 Can Differentiate to Definitive Endoderm

All hiPSC lines (WT, HS2ST1$^{+/-}$ and HS2ST1$^{-/-}$) were differentiated to DE in serum-free, GAG-free conditions on VTN-N and showed clear changes in morphology to a flatter endoderm cell type (**Figure 30**).

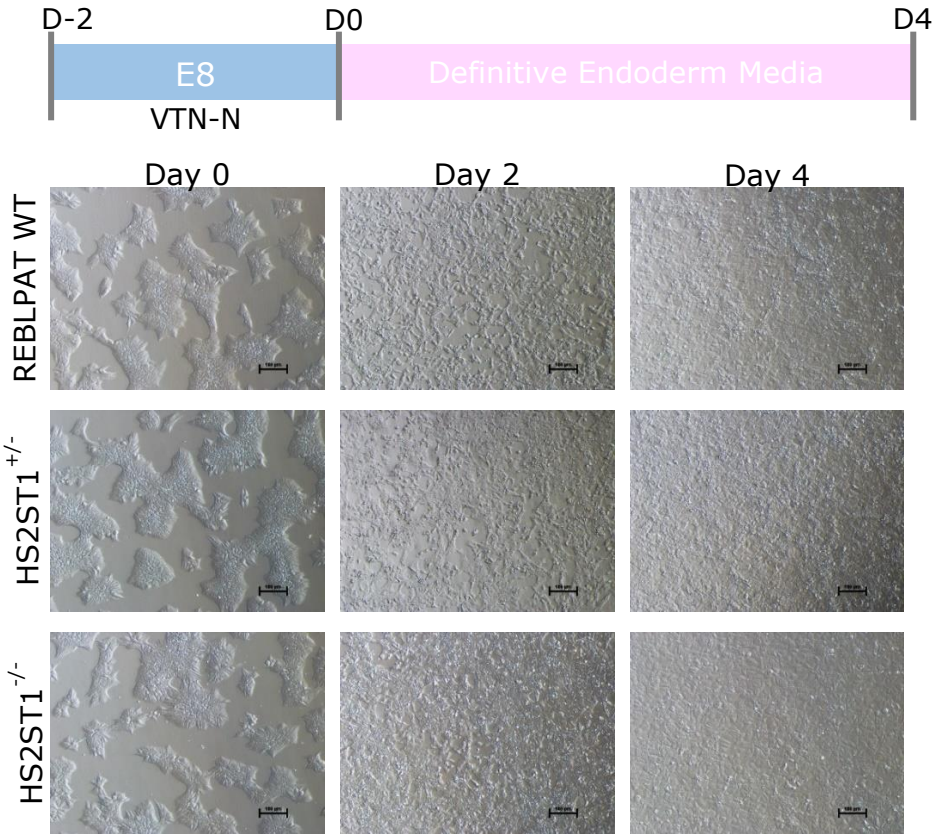


Figure 30 - Definitive Endoderm Differentiation. hiPSCs were seeded onto VTN-N at 20,000 cells/cm^2 and cultured for 4 days in DE medium. Scale bars 100 µm

All hiPSC lines displayed an increase in DE-associated genes (GATA4, SOX17, FOXA2, CXCR4) coupled with a decrease in pluripotency genes (NANOG, OCT4) as expected for cells undergoing active differentiation (**Figure 31A**). Interestingly, all hiPSC lines had similar gene expression of CXCR4, but GATA4 and SOX17 were dramatically increased in the HS2ST1$^{+/-}$ line when compared with the REBLPAT WT and HS2ST1$^{-/-}$ lines. Furthermore, HS2ST1$^{-/-}$ expressed decreased levels of FOXA2 (**Figure 31A**). The differentiation of all cell lines to DE was confirmed by flow cytometry, with all three hiPSC lines expressing similar levels of cKit and CXCR4 (**Figure 31B**).

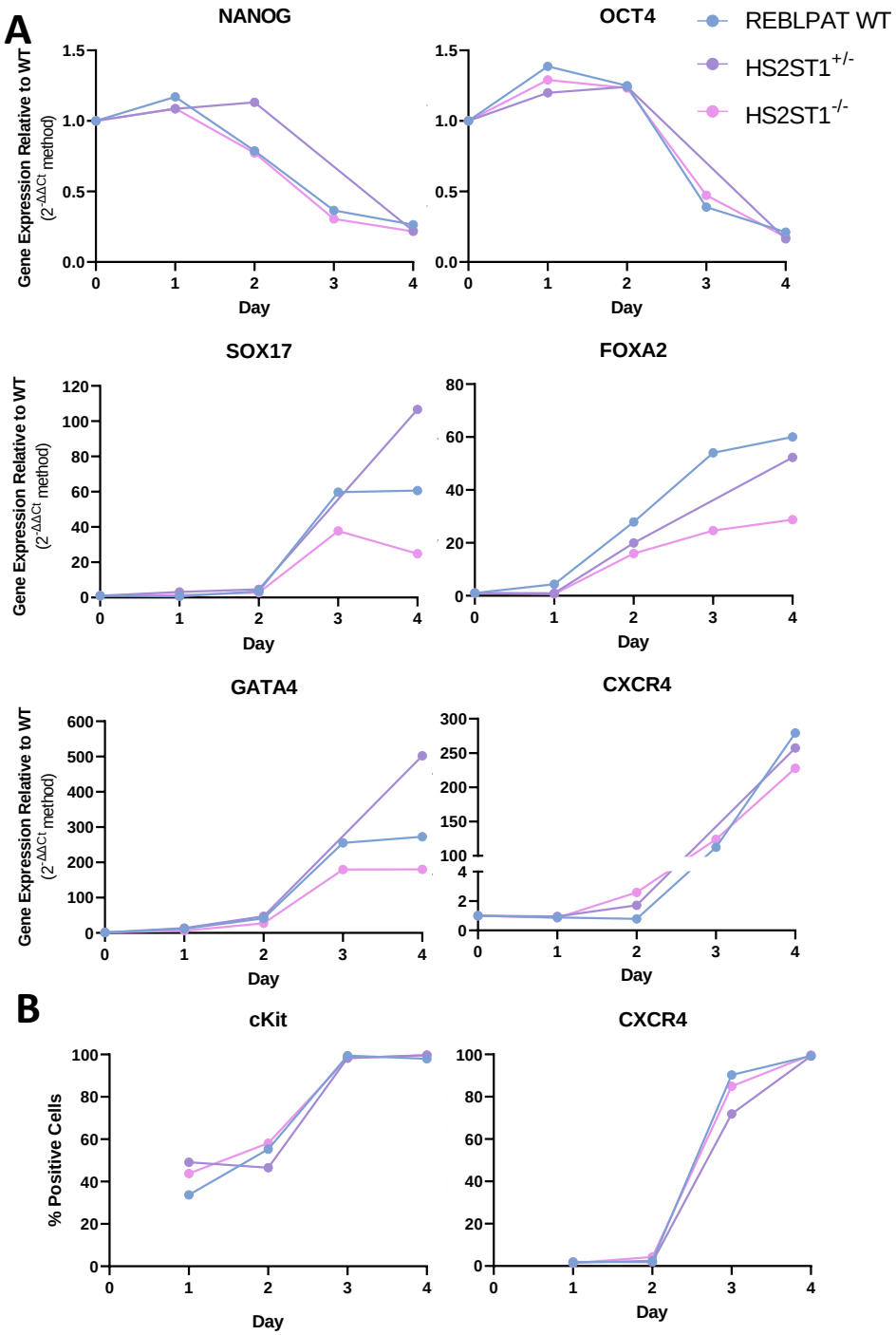


Figure 31 - hiPSCs with Altered HS2ST1 Function Differentiate to Definitive Endoderm. A) Gene expression of pluripotency genes (NANOG, OCT4) and endoderm genes (GATA4, SOX17, FOXA2, CXCR4). Expression normalised to PGBD. Error bars represent ± SEM from n=3 technical replicates (some error bars cannot be seen as they are smaller than the symbol). B) Percentage positive cells of endoderm markers (CXCR4, cKit) assessed by flow cytometry.

qRT-PCR results align with immunostaining images that also show a reduction in pluripotency markers (**Figure 32**) and an increase in endodermal markers (**Figure 33 and 34**).

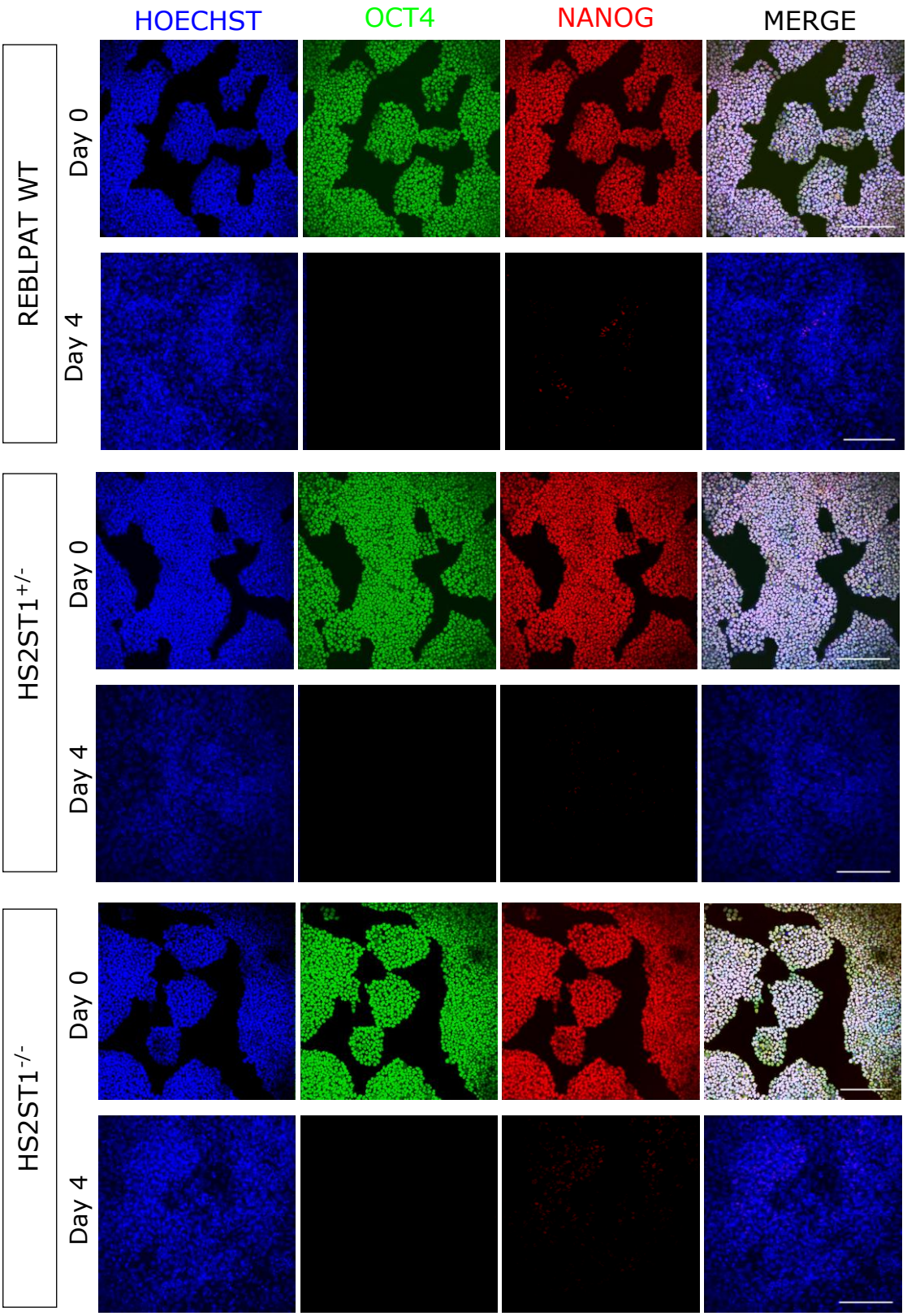


Figure 32 – Definitive Endoderm Differentiation (NANOG, OCT4; REBLPAT WT, HS2ST1$^{+/-}$, HS2ST1$^{-/-}$).

Immunofluorescence images of NANOG and OCT4 pluripotency markers on day 0 and day 4. Scale bars 200 μm

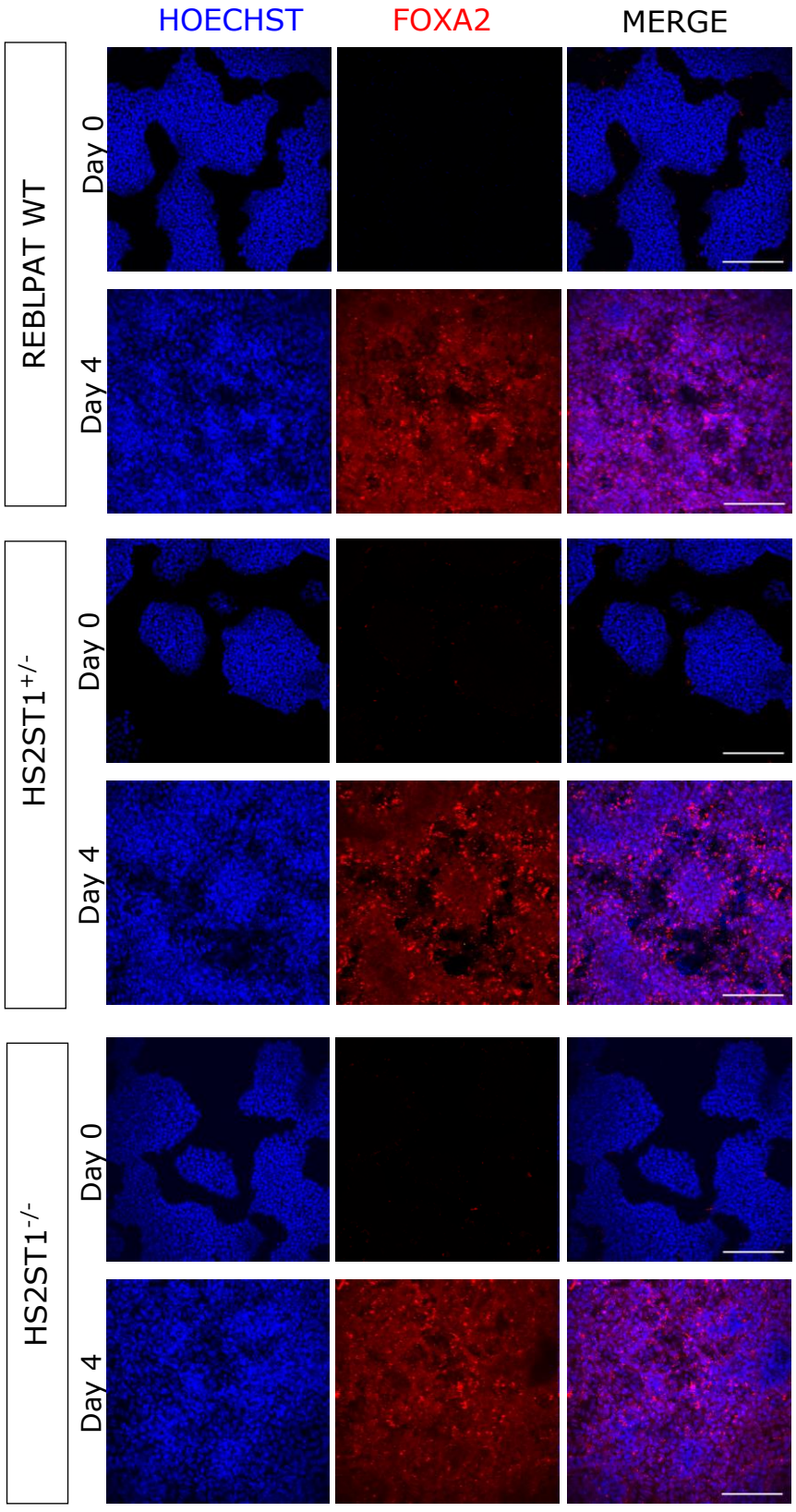


Figure 33 – Definitive Endoderm Differentiation (FOXA2; REBLPAT WT, HS2ST1$^{+/-}$, HS2ST1$^{-/-}$). Immunofluorescence images of FOXA2 endoderm marker on day 0 and day 4. Scale bars 200 µm

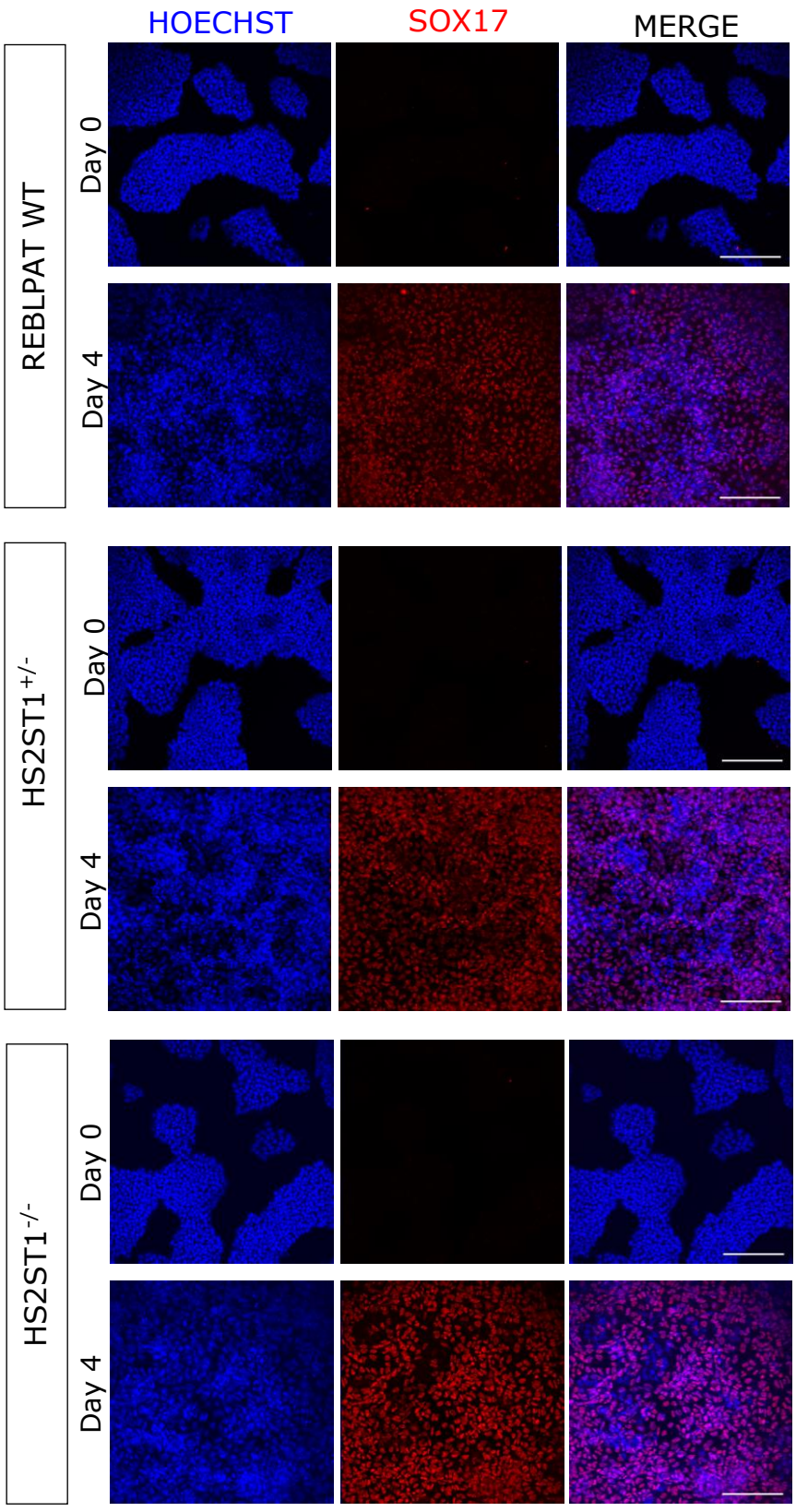


Figure 34 - Definitive Endoderm Differentiation (SOX17; REBLPAT WT, HS2ST1$^{+/-}$, HS2ST1$^{-/-}$). Immunofluorescence images of FOXA2 endoderm marker on day 0 and day 4. Scale bars 200 µm

5.3.3 hiPSCs with Functionally Altered HS2ST1 Can Differentiate to Neural Precursor Cells

The neural induction medium used for the NPC was purchased as neurobasal medium and neural induction supplement. As the supplement was proprietary, the precise composition was unknown. However, to ensure the supplement was GAG-free, HS compositional disaccharide analysis was run on the medium. The medium was found to contain no detectable HS disaccharides and could therefore be used without affecting interpretation of the impact of the knockout (**Appendix 3**).

Neural induction of hiPSCs was carried out for 7 days using commercially available PSC Neural Induction Medium. Initially, the protocol was tested on WT cells to optimize the seeding density. When seeded at lower densities, cells began to spontaneously differentiate and did not reach the desired morphology. When correctly formed, the individual NPCs look distinctly flatter when compared to hiPSCs and, by day 7, the cells form an uneven, dense layer of cells with an undulating appearance. However, when seeded too high, the cells detached from the plate (**Figure 35**). When REBLPAT WT hiPSCs were seeded at 27,000 cells/cm^2, the cells reach an NPC-like morphology and remained attached to the VTN-N and could be collected for analysis. qRT-PCR results showed a clear reduction in pluripotency markers (NANOG, OCT4) and an increase in NPC markers (NESTIN, SOX1) (**Figure 36**).

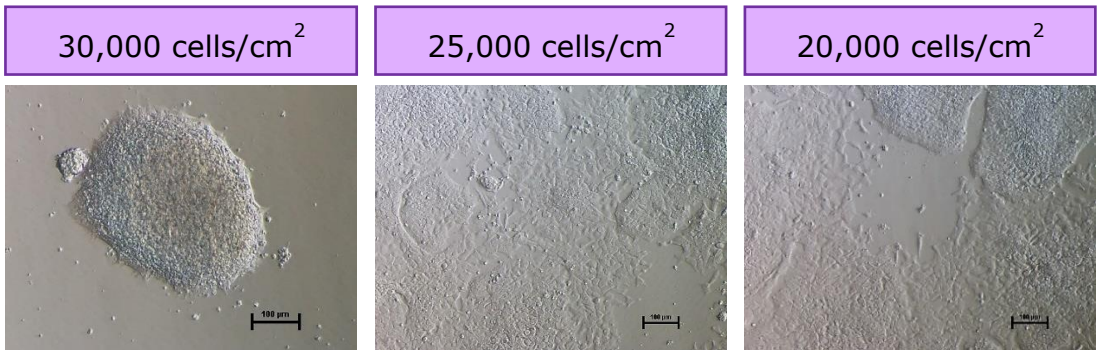


Figure 35 - Optimizing Seeding Densities for Neural Precursor Cell Differentiation. Seeding cells at higher density results in cell detachment while seeding at lower densities results in non-neural differentiation. Optimization carried out using REBLPAT WT cells. Images taken at day 6. Scale bars 100 µm

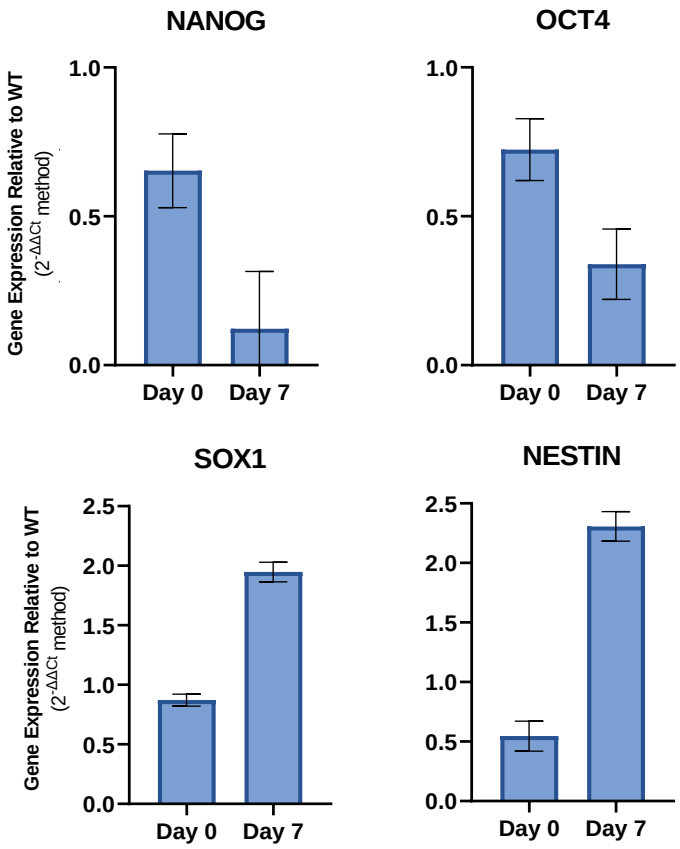


Figure 36 - Neural Progenitor Cells Form in REBLPAT WT hiPSC After 7 Days. qRT-PCR results of NPC differentiation of REBLPAT WT cells seeded at 27,500 cells/cm^2. Gene expression was normalized to PBGD. Error bars represent ± SEM from n=3 technical replicates. N=1 individual experiment.

When seeded at this 27,500 cells/cm^2, all three hiPSC lines show clear changes in morphology as they progress from pluripotent stem cells to NPCs as expected. The morphology, by day 7, can be described as a layer of NPCs covering the surface, that is 3D in some areas giving an uneven appearance in the plate. The images show the cells formed an NPC like morphology and both HS2ST1$^{+/-}$ and HS2ST1$^{-/-}$ hiPSCs show a similar phenotype to the WT cell line (**Figure 37**).

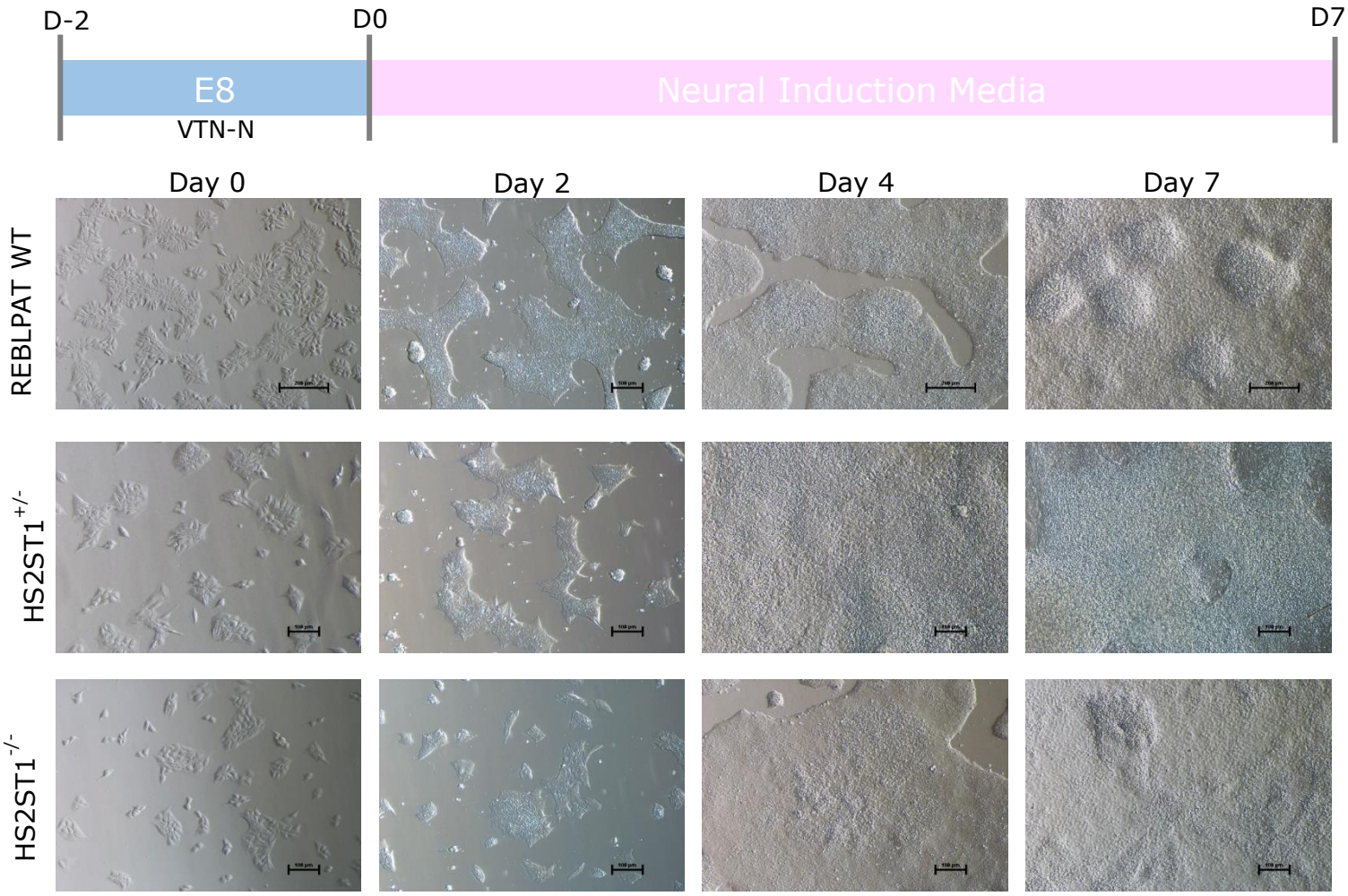


Figure 37 - Neural Precursor Cell Differentiation. hiPSCs were seeded onto VTN-N and cultured for 7 days in Neural Induction Medium. Morphology changes to densely packed, 3D structures. Scale bars 100 µm

Additionally, immunostaining revealed a total loss of OCT4 and NANOG by day 7. However, unlike NANOG, OCT4 was still detectable around day 4 in all three hiPSC lines (**Figure 38 and 39**). In REBLPAT WT and HS2ST1$^{+/-}$ lines, SOX2 is present in cells at day 0, reduced at day 4 and detectable again by day 7. As expected, NESTIN (an NPC marker), is not present at day 0, but the signal increases from day 4 onwards (**Figure 40**). In HS2ST1$^{-/-}$ hiPSCs, NESTIN staining in HS2ST1$^{-/-}$ is similar to the other two cells lines, undetectable at day 0 but present at days 4 and day 7 (**Figure 41**). PAX6 (also an NPC marker) staining is not present at day 0 in all three hiPSCs lines but is detectable at very low levels at day 4 with a much stronger signal at day 7. It is important to note that PAX6 staining at day 7 in HS2ST1$^{-/-}$ appears decreased when compared with REBLPAT WT and HS2ST1$^{+/-}$ at day 7 (**Figure 42 and 43**).

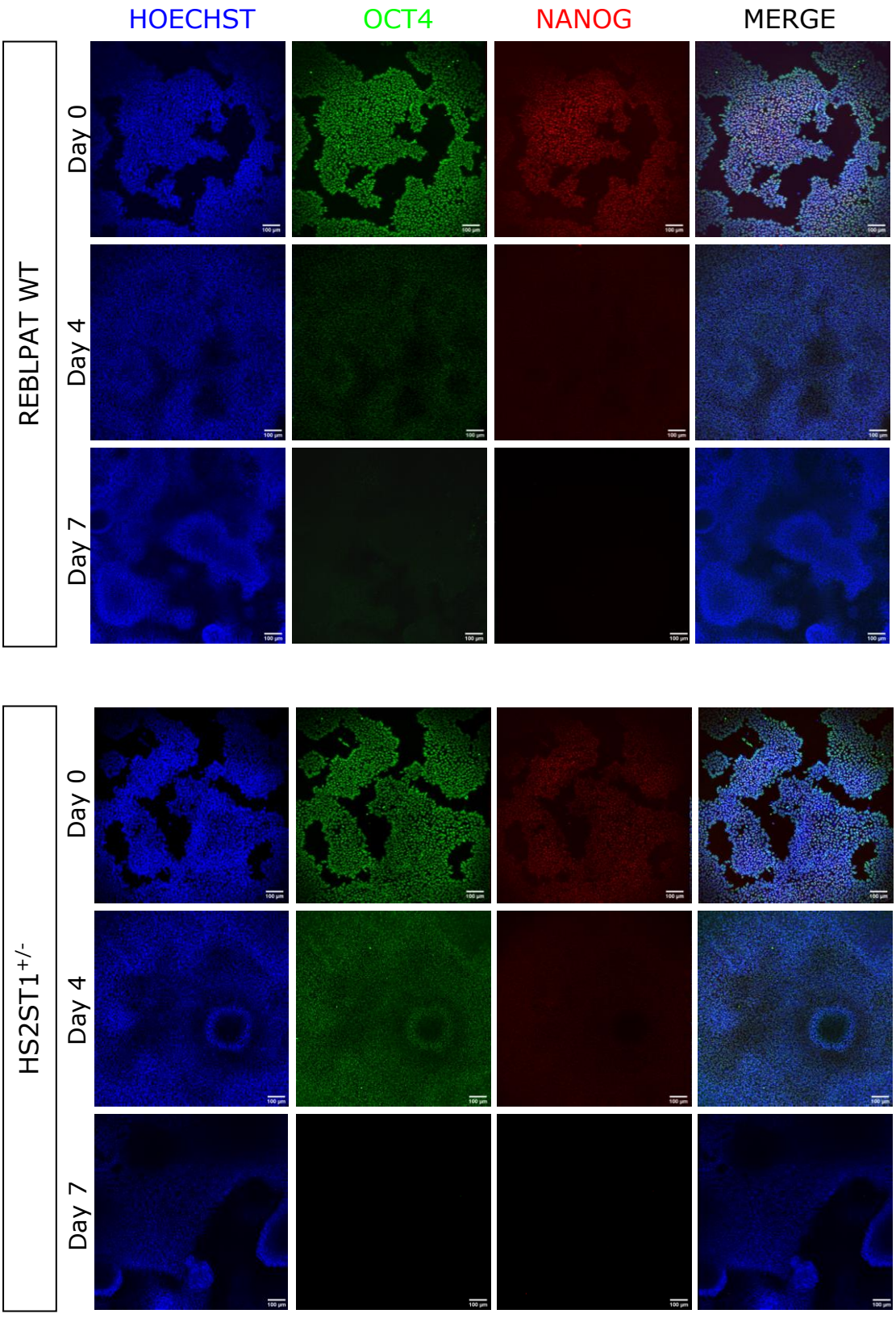


Figure 38 – Neural Precursor Cell Differentiation (NANOG, OCT4; REBLPAT WT, HS2ST1$^{+/-}$). Immunofluorescence images of NANOG and OCT4 pluripotency markers on day 0, day 4 and day 7. Scale bars 100 µm

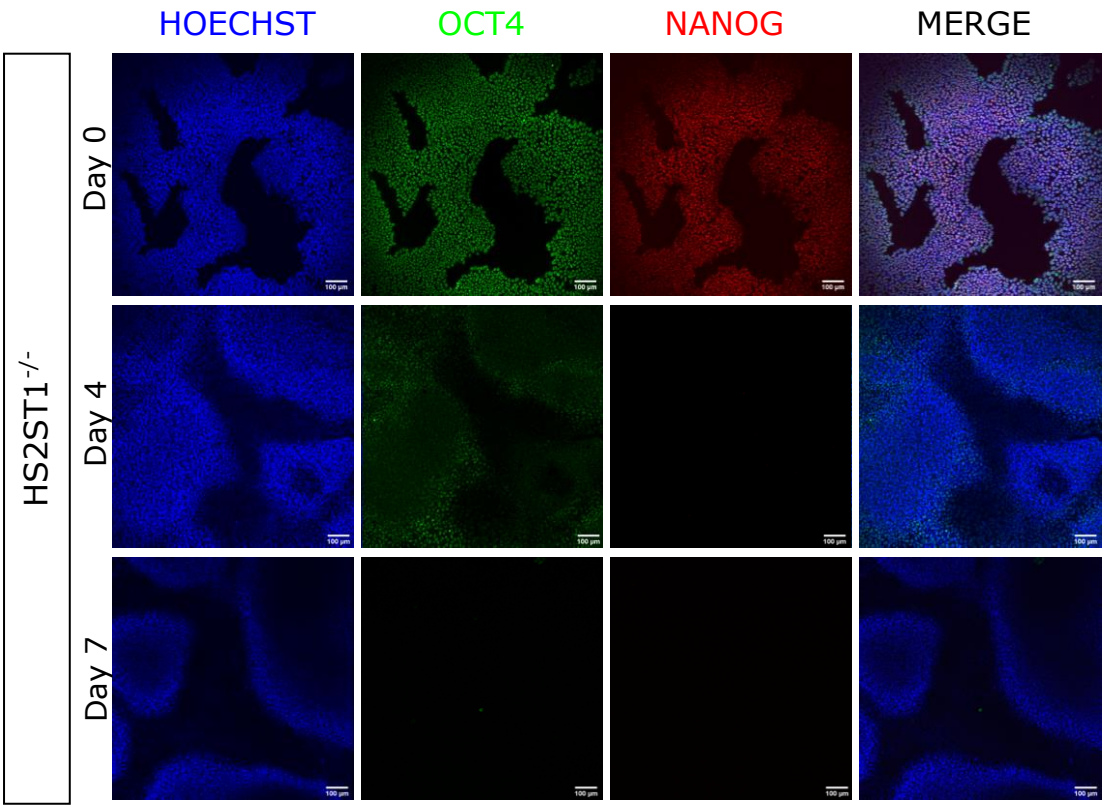


Figure 39 – Neural Precursor Cell Differentiation (NANOG, OCT4; HS2ST1$^{-/-}$). Immunofluorescence images of NANOG and OCT4 pluripotency markers on day 0, day 4 and day 7. Scale bars 100 µm

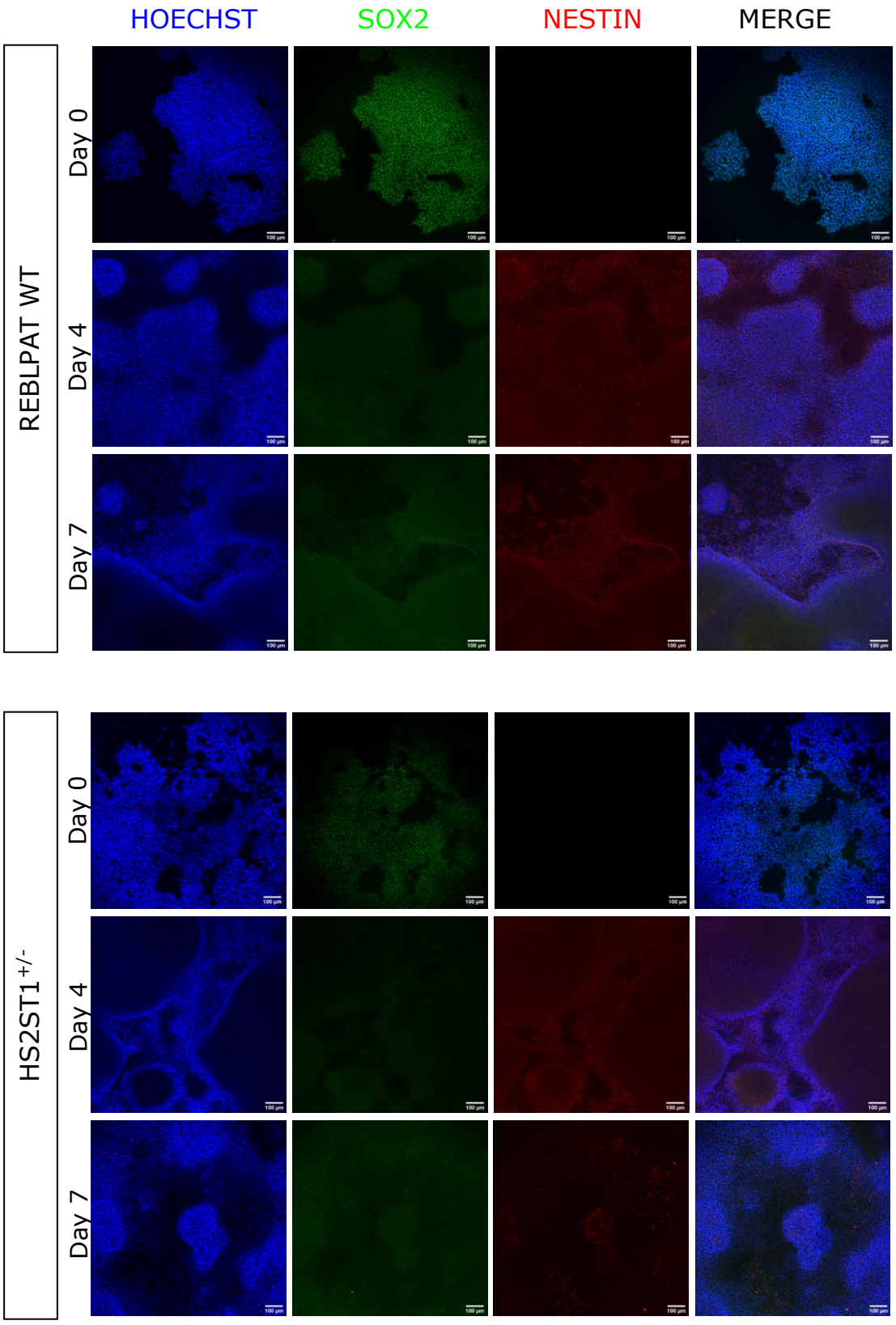


Figure 40 – Neural Precursor Cell Differentiation (NESTIN, SOX2; REBLPAT WT, HS2ST1$^{+/-}$). Immunofluorescence images of NESTIN (NPC marker) and SOX2 (NPC and pluripotency marker) on day 0, day 4 and day 7. Scale bars 100 µm

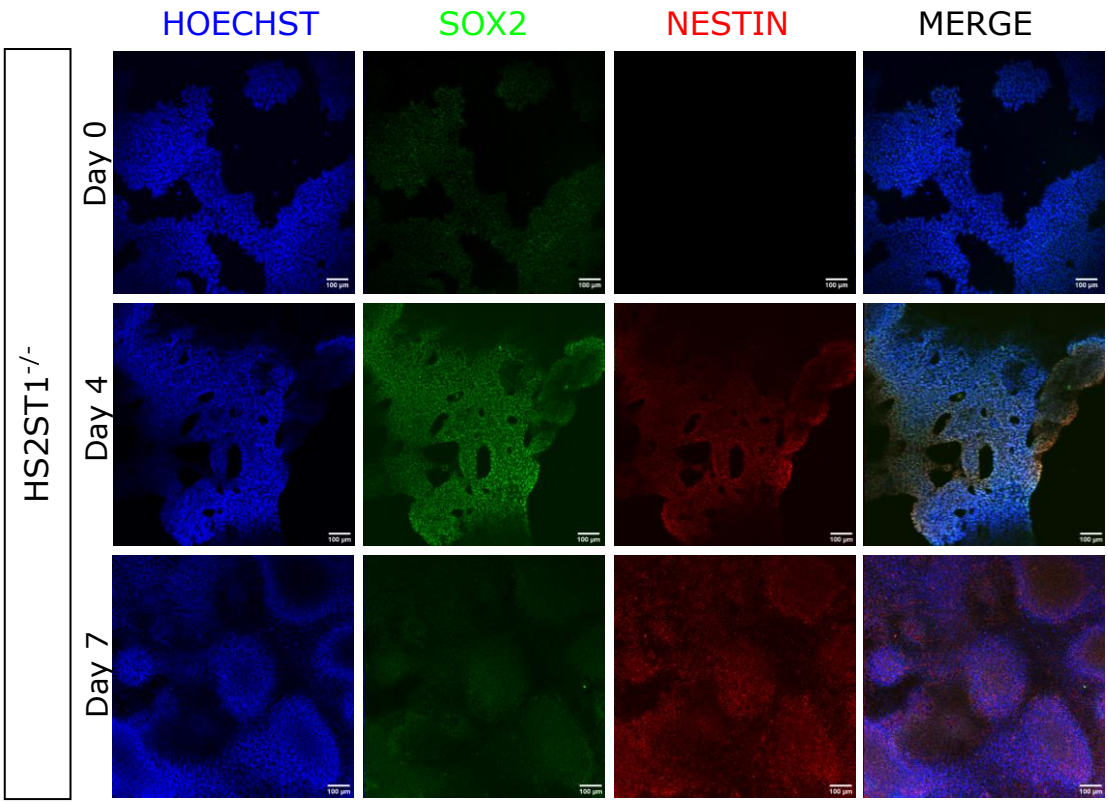


Figure 41 – Neural Precursor Cell Differentiation (NESTIN, SOX2; HS2ST1$^{-/-}$). Immunofluorescence images of NESTIN (NPC marker) and SOX2 (NPC and pluripotency marker) on day 0, day 4 and day 7. Scale bars 100 µm

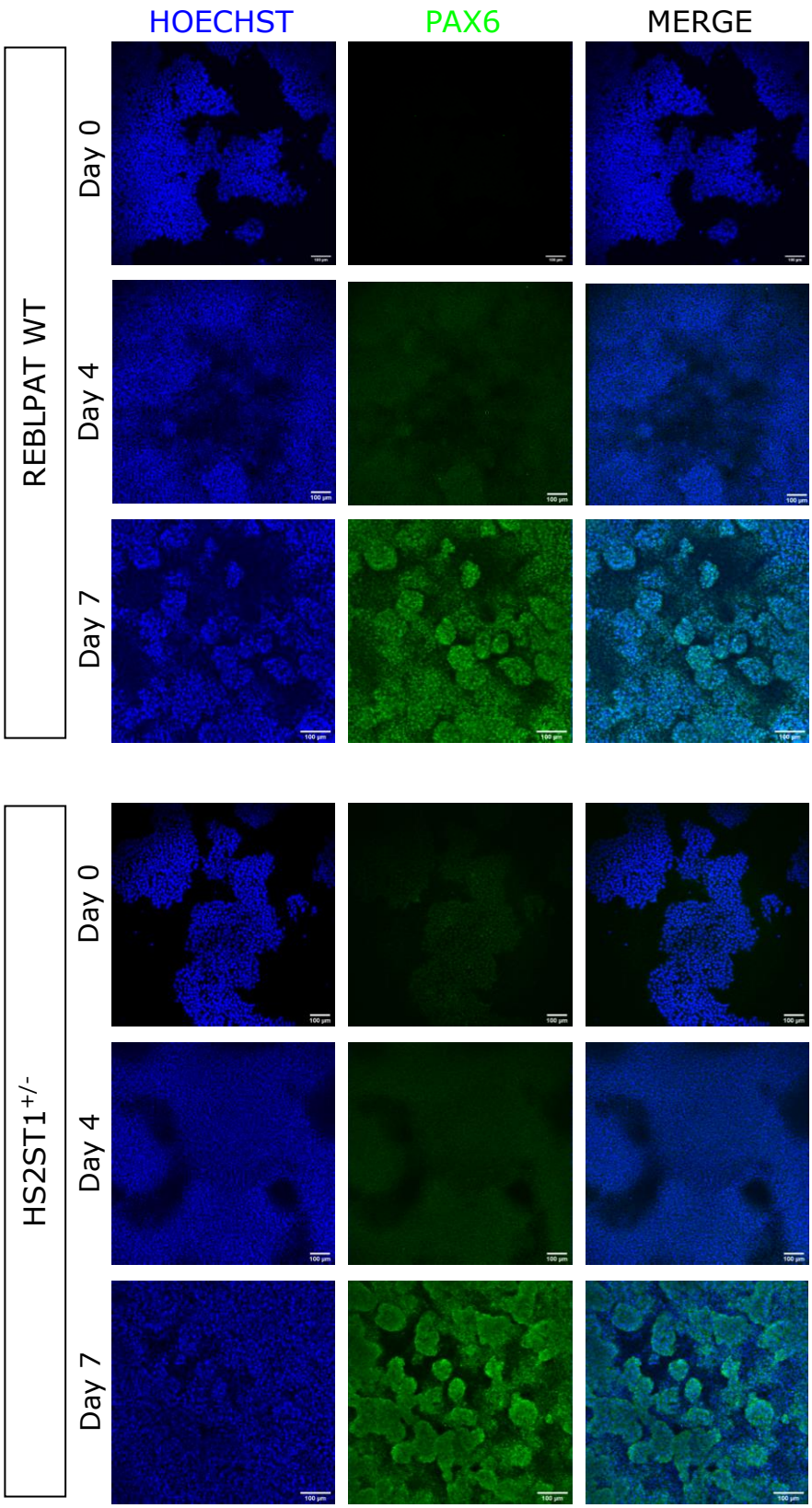


Figure 42 – Neural Precursor Cell Differentiation (PAX6; REBLPAT WT, HS2ST1$^{+/-}$). Immunofluorescence images of PAX6 NPC marker on day 0, day 4 and day 7. Scale bars 100 µm

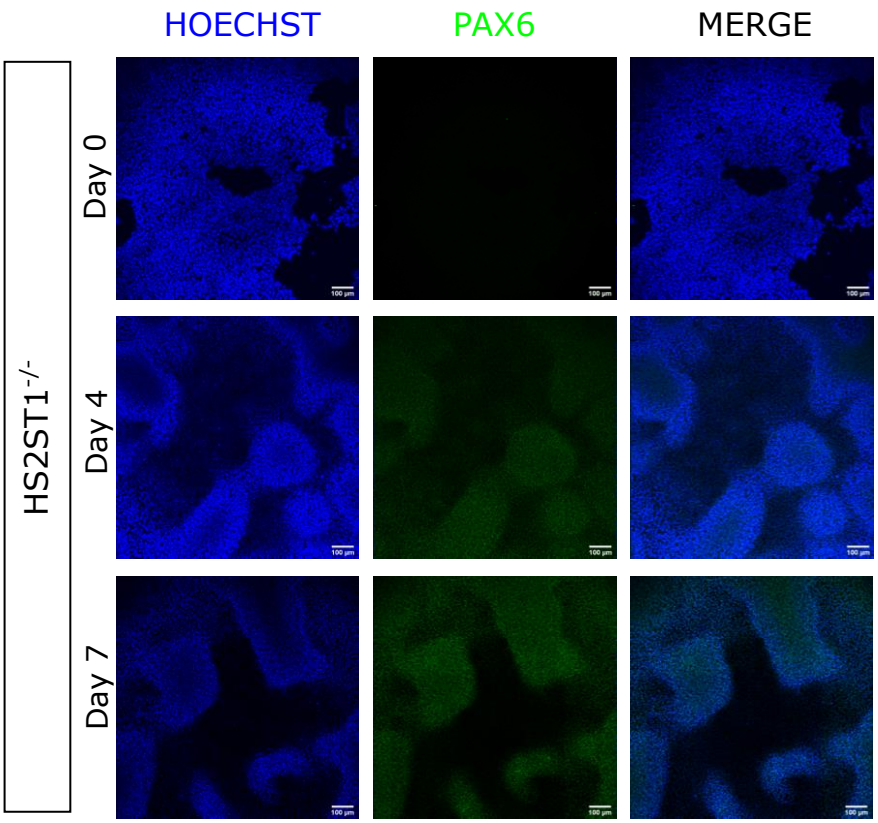


Figure 43 – Neural Precursor Cell Differentiation (PAX6; HS2ST1$^{-/-}$). Immunofluorescence images of PAX6 (NPC marker) on day 0, day 4 and day 7. Scale bars 100 µm

5.3.4 Optimizing A Protocol for Neural Differentiation

Following neural induction, NPCs were split and reseeded and maintained in N2B27 for 14 days to stimulate neuron formation. Although qRT-PCR shows a reduction of pluripotency makers, neural markers are also downregulated. In all three hiPSC lines, SOX1 appeared upregulated on day 7, with the largest increase (14-fold) seen in HS2ST1$^{+/-}$. By day 21, SOX1 was still highly upregulated in HS2ST1$^{+/-}$ while only marginally upregulated in HS2ST1$^{-/-}$ and downregulated in REBLPAT WT cells. NESTIN was increased 3-fold in HS2ST1$^{-/-}$ at day 7 and 2-fold at days 14 and 21, while the REBLPAT WT and HS2ST1$^{+/-}$ cells showed NESTIN to decrease throughout the course of the differentiation. PAX6 and β-III-TUBULIN were downregulated in all three hiPSC lines across the full 21 days (**Figure 44**). It is important to note that this differentiation was only carried out once.

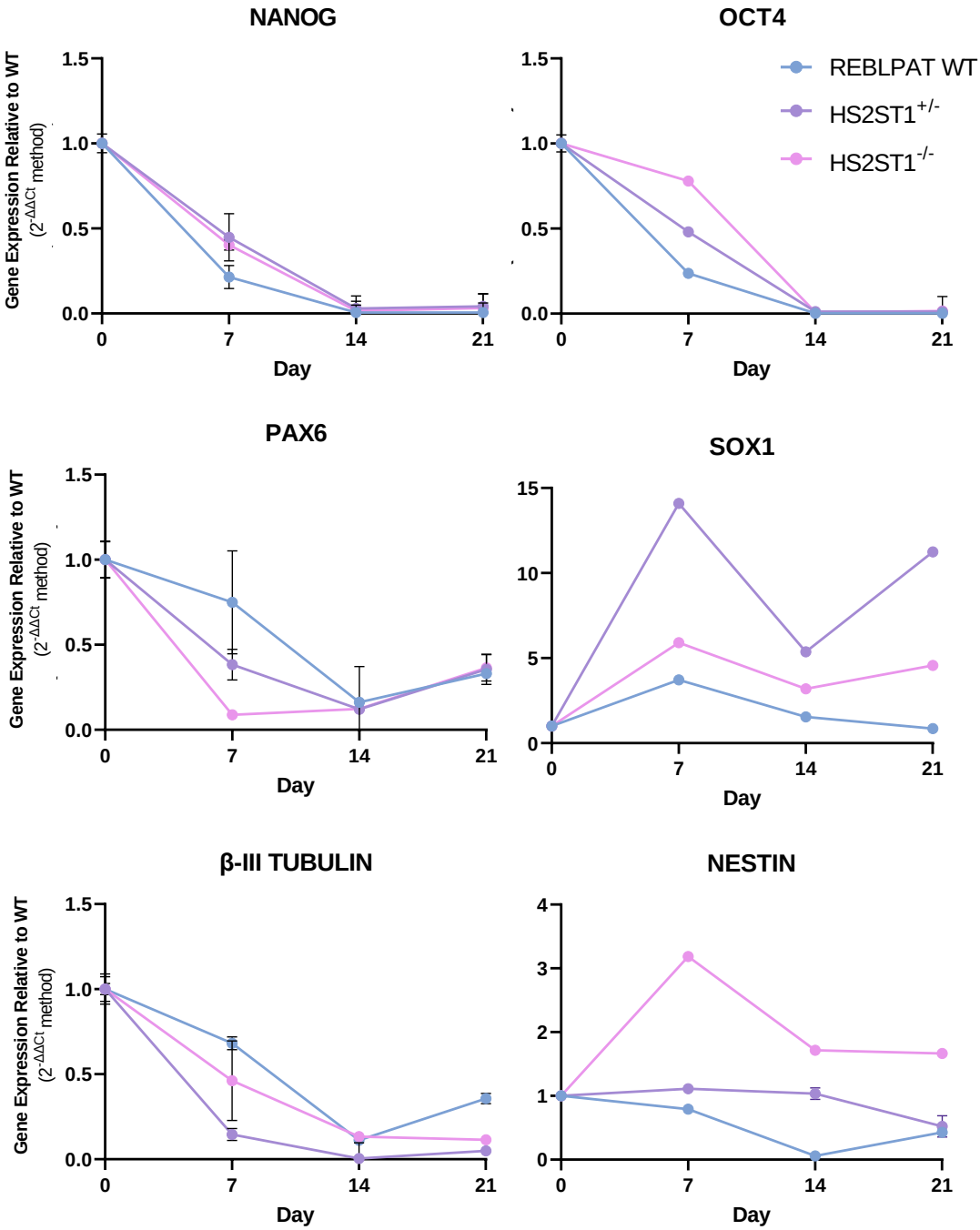


Figure 44 - Neural Differentiation. Gene expression of pluripotency markers (NANOG, OCT4), neural precursor markers (PAX6, SOX1, NESTIN) and a neural marker (β-III-TUBULIN). Expression was normalized to PBGD. Error bars represent ± SEM from n=3 technical replicates. N=1 individual experiment

5.3.5 HS2ST1$^{-/-}$ hiPSCs Display Reduced FGFR3 Levels

As discussed in Chapter 1, HS is known to interact with FGF2 to help maintain pluripotency and drive differentiation (Turnbull *et al.*, 1992; Mossahebi-Mohammadi *et al.*, 2020). Additionally, 2-O-desulphated heparin was unable to bind to FGF2 indicating this modification is essential for the interaction (Guimond *et al.*, 1993; Maccarana, Casu and Lindahl, 1993). As hiPSCs with altered HS2ST1 activity still appear to maintain pluripotency and are able to differentiate, we proposed to assess the FGF2 signalling capacity. The initial experiment was designed to assess signalling by treating cells with differing levels of FGF2 and quantifying levels of ERK1/2 phosphorylation via western blot. Unfortunately, this was not possible during the time frame of this project. However, to determine whether there are any differences in FGFR expression in functional knockouts, qRT-PCR of hiPSCs cultured under standard pluripotency conditions was used. No significant increase in expression of the FGF receptors (FGFRs) was seen in hiPSCs with altered HS2ST1 function (HS2ST1$^{+/-}$) when compared with wild type hiPSCs. However, a significant decrease of FGFR3 expression was observed in HS2ST1$^{-/-}$ hiPSCs when compared with the relevant control (**Figure 45**).

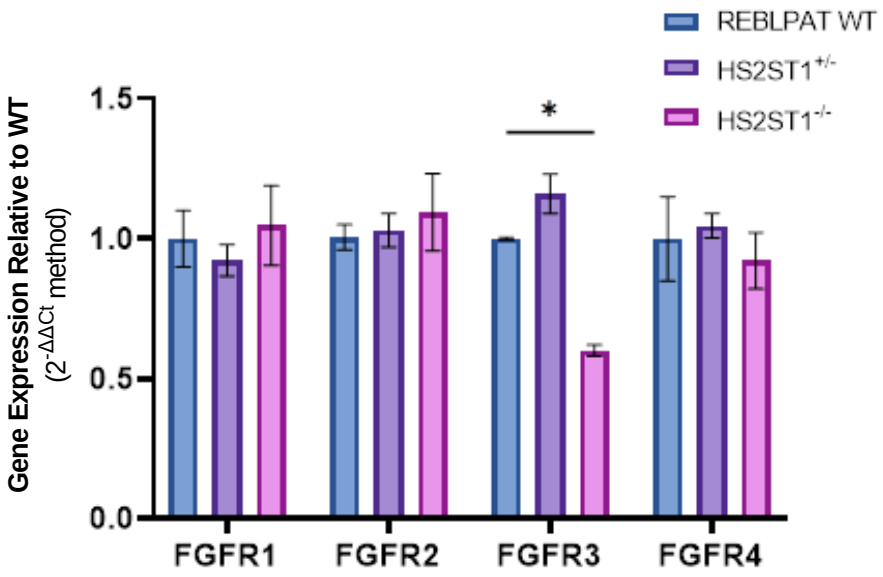


Figure 45 - HS2ST1$^{-/-}$ hiPSCs Show Reduced FGFR3 Expression. qRT-PCR showing expression of FGFRs in WT, HS2ST1$^{+/-}$ and HS2ST1$^{-/-}$ hiPSCs cultured in standard pluripotency conditions (E8 and VTN-N coating). Expression was normalized to PBGD. Error bars represent ± SEM from N=3 replicates. Tukey's multiple comparison test was used to calculate statistically significant differences. *p<0.05.

5.4 Discussion

5.4.1 Altered Heparan Sulphate Structure Does Not Impair Definitive Endoderm and Neural Precursor Cell Formation

Definitive endoderm formation is characterized by the presence of SOX17 and FOXA2 as well as a double positive population when stained with CXCR4 and cKit by assessment using flow cytometry (Takata *et al.*, 2011; Hou *et al.*, 2018; Sahabian *et al.*, 2021). In the experiments presented here, these criteria were met by cells defective in HS2ST1 activity. Therefore, from the data presented in this chapter it can be concluded that loss of HS2ST1 does not impair DE differentiation as functional knockouts appeared no different to WT. This aligns with previous work carried out in HS deficient hiPSCs (*EXT2*$^{-/-}$) which were also able to differentiate to DE (Jamie Thompson, personal communication). In contrast, FOXA2 and SOX17 were vastly upregulated (300- and 300,000-fold respectively) by day 2 in *EXT2*$^{-/-}$ hiPSCs when compared with WT control when assessed by qRT-PCR. Additionally, in HS-deficient hiPSCs cKit is shown to be positive in 80% of KO cells by day 1 compared to around 50% in REBLPAT WT hiPSCs. This study makes a tentative conclusion that hiPSCs deficient in HS were able to differentiate to DE more rapidly than hiPSCs with WT HS, though the differentiation was only carried out once (Jamie Thompson, personal communication).

NPC formation was originally tested in REBLPAT WT cells as the protocol had not previously been used at the University of Nottingham so initially seeding density was optimized and the protocol tested in WT hiPSCs. When HS2ST1$^{+/-}$ and HS2ST1$^{-/-}$ were differentiated alongside REBLPAT WT hiPSCs, all lines appeared to form an NPC-like morphology at similar rates. NPC formation was confirmed by the presence of PAX6, SOX2 and NESTIN. NESTIN is widely used as a marker for NPCs as it is very specifically expressed in these cells and not in more mature neurons (Hu *et al.*, 2016). PAX6 and SOX2 are also recognised NPC markers that are known to interact with each other (Wen *et al.*, 2008). PAX6 and NESTIN control the proliferation and differentiation of NPCs while SOX2 helps to maintain their self-renewal. Although staining appears weak, NESTIN is present at day 4 and day 7 in all hiPSC lines. SOX2 appears to be present across the whole differentiation timeline at varying intensities. This is because SOX2 is also a

pluripotency marker, so this is to be expected (Feng and Wen, 2015). As with the DE differentiation, HS-deficient hiPSCs were able to form NPCs at a faster rate when compared to WT, but with varying success (Jamie Thompson, personal communication). Although the findings presented here align with other hiPSC models with no, or abnormal, HS it is still surprising that differentiation is not altered from wild type, especially considering the essential roles of HS-dependent signalling pathways involved in the differentiation to these lineages.

After 7 days, the NPCs were plated out onto recombinant laminin-111 and cultured in N2B27 to differentiate the NPCs to neurons. It was hypothesized that PAX6, NESTIN and SOX1 would be expressed early in the differentiation, before β-III-TUBULIN expression increased as immature neurons were formed. However, due to the downregulation of neural markers, it was concluded that the differentiation was unsuccessful. Seeding density is extremely important in neural differentiations, as noted with the seeding density of hiPSCs when forming NPCs in the present study. Optimization of differentiation protocols often takes a long time and to create a protocol with high replicability is difficult. As this is the first time this differentiation has been performed, a thorough optimization of the process is recommended. Once the neural protocol is able to be successfully repeated in REBLPAT WT hiPSCs, the functional knockouts should be tested. It is important to note that neural differentiations are often long protocols with low reproducibility and low efficiency so it may not be a feature of the knockouts, but more the differentiation being attempted (Hu *et al.*, 2010; Zhang *et al.*, 2013; Frega *et al.*, 2017). It would be difficult to mitigate these issues, especially in a GAG-free environment where cells are likely to be more stressed, but development of a robust protocol remains a priority.

Differentiating HS2ST1$^{+/-}$ and HS2ST1$^{-/-}$ hiPSCs to a mesoderm lineage is important to determine whether there are any defects with specific lineages. HS deficient hiPSCs were differentiated to beating cardiomyocytes for 18 days alongside a WT control with similar results (Jamie Thompson and James Smith (UEA), personal communication). However, when the same cells were differentiated down the cardiac lineage for longer, differences in the

differentiation capability became clear (James Smith, personal communication). Therefore, it may only be in more extensive differentiation protocols that differences between WT hiPSCs and hiPSCs with altered HS become clear. For this reason, extending the endoderm differentiation beyond 4 days would be recommended.

5.4.2 HS2ST1$^{-/-}$ hiPSCs Display Decreased Expression of FGFR3

FGFR3 expression was significantly decreased in HS2ST1$^{-/-}$ hiPSCs (n=3). Although it is unknown whether protein levels are also decreased, it could still provide some clinical significance. *FGFR3* is a gene encoding one of the FGF receptors, a member of the receptor tyrosine kinase family. FGFRs are tissue specific and individual isoforms have different specificity for specific FGFs in distinct tissues (Xie *et al.*, 2020). FGFR3 is known to negatively regulate endochondral ossification by inhibiting proliferation and differentiation of chondrocytes (Wang *et al.*, 2015). Furthermore, FGFR3 is known to be expressed in osteoblasts, the cells responsible for bone matrix deposition (Caetano-Lopes, Canhão and Fonseca, 2007; Xie *et al.*, 2020).

Fgfr3 knockout mice have skeletal abnormalities that include scoliosis and tail distortions (Deng *et al.*, 1996). More specifically, in chondrocytes with an *Fgfr3* deficiency, osteoblasts display increased mineralization (Wen *et al.*, 2016). As *FGFR3* expression is reduced in HS2ST1$^{-/-}$ hiPSCs, it can be hypothesized that differentiation to osteoblasts and chondrocytes may be altered from wild type. Additionally, if the decrease in *FGFR3* expression is due to the HS2ST1 knockout, it can be hypothesized that the same decrease would been seen in human patients with alterations in HS2ST1 as well as mouse models, although this has not been investigated. However, if the FGFR3 levels are reduced, it could provide an insight into the skeletal defects seen across both species. *Hs2st1* null mice display increased bone ossification which could align with a reduction of FGFR3 function in the chondrocytes. The short stature seen in patient 1 could also be as a result of FGFR3 abnormalities, as mutations in the gene are known to cause hypochondroplasia and achondroplasia due to a reduction in the rate of cartilage and bone formation from the growth plate (Wang *et al.*, 2015). Furthermore, it has also been reported that partial

loss of FGFR3 function results in a taller stature and camptodactyly, symptoms noted in patients 2 and 4 presented in this study (Toydemir *et al.*, 2006).

For a more detailed investigation, assessment of the *FGFR* expression in *Hs2st1*$^{-/-}$ mouse embryonic stem cells and patient fibroblasts would be useful. This would provide insight into whether a reduction in FGFR3 contributes to the skeletal phenotype seen in both species. Additionally, investigating whether FGFR protein levels are altered is also important. Furthermore, assessing the FGF signalling capacity, specifically FGF2, of the hiPSCs would also be necessary, as previously it has been seen that hiPSCs are still able to elicit a response to FGF2 in HS deficient cell lines (Jamie Thompson, personal communication). Therefore, it can be hypothesised that the HS2ST1 deficient hiPSCs will also still be able to signal via FGF2.

5.4.3 How Are hiPSCs With Functionally Altered HS2ST1 Maintaining Pluripotency and Differentiating?

A previously discussed, the IdoA(2S) disaccharide found in HS is key for high affinity binding of FGF2, and FGF2 is necessary to maintain pluripotency (Turnbull *et al.*, 1992; Maccarana, Casu and Lindahl, 1993; Vallier, Alexander and Pedersen, 2005). Furthermore, many other signalling factors that rely on HS interaction are required for both pluripotency and differentiation including Wnt, TGF β and BMPs as shown in **Figure 10 and 11** (Kuo, Digman and Lander, 2010; Mii and Takada, 2020). Therefore, it is striking that these hiPSCs express pluripotency factors and are able to differentiate to DE and NPCs.

Models that have altered HS display enzymatic compensation. For example, in the present model, loss of 2-O-sulphation is compensated for by an increased in 6-O-sulphation and N-sulphation. Furthermore, removal of Hs6st1 and Hs6st2 results in increased 2-O-sulphation in HS of mice (Sugaya et al., 2008). In hiPSCs with a total loss of HS, the ECM still contains CS/DS, but is also shown to be KS rich. The increase in KS is likely to be compensatory and it was hypothesized that this increase allows for continued FGF2 signalling (Jamie Thompson, personal

communication). Additionally, the HS deficient hiPSCs were still able to differentiate in the same way as wild type controls, despite the apparent need for HS to mediate essential signalling pathways, further indicating the compensatory effects of KS (Jamie Thompson, personal communication). It is difficult to determine whether the compensatory increase in 6S and NS in HS2ST1-/- hiPSCs is enough to allow the cells to differentiate and maintain pluripotency. The increase in sulphation may be necessary to maintain some of the functions of HS, but an increase in other sulphated GAGs may also occur.

5.4.4 Summary of Characterization of hiPSCs With Altered HS2ST1

In conclusion, it has been presented here that despite lacking the HS 2-O-sulphation modification, hiPSCs are still able to maintain pluripotency and differentiate to DE and NPCs. Unfortunately, neural differentiation requires further optimisation and could not be completed as part of this study. A particularly interesting finding was the reduction of FGFR3 expression. To investigate whether $HS2ST1^{-/-}$ and $HS2ST1^{+/-}$ hiPSCs have an altered ability to form bone cells, such as osteoblasts and osteoclasts, an *in vitro* differentiation would need to be carried out. Furthermore, due to the penetrant kidney phenotype seen in the mouse models and some of the patients, a kidney differentiation would also be of interest.

Chapter 6

Final Conclusions

6.1 Final Discussion and Conclusions

This doctoral thesis sheds light on the complexities of GAG structure and provides additional insight as to how the HS biosynthetic machinery is interlinked. Despite considerable effort over many years, including recent breakthroughs in structural analyses, many of the specific interactions of GAG biosynthetic enzymes remain unknown. However, it is clear that when alterations in the biosynthetic machinery arise, compensatory increases in activity of the other modification enzymes can occur, impacting on the observed structure of GAGs synthesised by the cells. This is also dependent on the specific biosynthetic enzymes involved as in some situations, for example when NDST1 activity is lost, sulphation is not compensated for by other sulphotransferases and the HS chains remain under sulphated. Additionally, the role of GAGs and their interactions with essential growth factors, particularly in complex biological situations where multiple interacting pathways are in play, remains somewhat of a mystery. Based on previous published research, it could be said that the data in this thesis indicates that, despite the loss of an essential HS modification, other sulphotransferase enzymes can act in a compensatory manner to maintain fundamental cellular process. However, as *EXT2*$^{-/-}$ hiPSCs still appear to maintain fundamental cellular functions despite producing no HS, biosynthetic enzyme compensation may not be the reason for this. It is important to note that HS deficient hiPSCs did produce increased levels of other sulphated GAGs, in particular KS, which may be the reason they are still able to act in a similar manner to WT hiPSCs (Jamie Thompson, personal communication).

Despite the termination of work due to the COVID-19 pandemic, some conclusions are still able to be drawn from Chapter 3. The chapter details using an HDR method of DNA delivery of CRISPR/Cas9. The main issue with the work was the generation of the plasmids. Therefore, following the change in focus of the PhD project, this issue was taken into consideration. When the CRISPR/Cas9 of *HS2ST1* was planned, it was decided against using

DNA delivery of the machinery, due to the issues faced previously. It was decided that RNP delivery of CRISPR/Cas9 would be used, which had great success. It would be interesting to retarget the *SGSH* gene using RNP delivery to determine whether plasmid generation was the only issue with this part of the project, or whether the gene itself is intrinsically hard to target.

The results presented in chapter 4 initially describe the HS composition of the first patients discovered to have mutations in the *HS2ST1* gene. From this, chapter 4 goes on to discuss the generation, and subsequent KO validation of HS2ST1 deficient hiPSCs. This is the first report of making HS2ST1 deficient hiPSCs, likely due to the fact that it was thought *HS2ST1* mutations had no clinical relevance as there were no known human patients. Early work focused on the generation of mutant CHO cells, mice and mESCs, and these models provided insight into how HS2ST1 was involved in fundamental developmental pathways (Bai and Esko, 1996; Bullock *et al.*, 1998; Merry *et al.*, 2001). However, these models were unable to provide a controlled environment in which different mutations could be investigated and how each affects a multitude of differentiation and signalling pathways. Furthermore, fundamental differences between CHO cells, mice and humans means that conclusions drawn from mESC investigations may not be applicable to human developmental pathways.

Although one of the major findings of chapter 4 was that, contrary to popular belief, humans were able to survive without the HS 2-O-sulphation modification, the generation of the HS2ST1$^{-/-}$ hiPSCs is also noteworthy. The lack of antibodies able to detect endogenous levels of the enzyme meant that an ScFv antibody with a preference for the 2-O-sulphation motif was used instead. However, as this is only a preferential epitope, the antibody was used in two validation experiments – immunostaining and flow cytometry. In addition, the HS composition was investigated using depolymerisation to disaccharides and detection using AMAC labelling and RP-HPLC. Although each of these experiments was only carried out once, the three different techniques together provide the validation needed for knockout confirmation. Furthermore, the different methods also provide extra assurance as AO4B08 only binds 2-O-sulphated regions preferentially, and non-

exclusively. Therefore, the RP-HPLC provides an antibody-free detection method to mitigate issues surrounding alternative binding. However, the reverse is still true, if the 2-O-sulphated disaccharides are below the detection limit of the HPLC, they will have been picked up using AO4B08 via immunostaining, or flow cytometry. Using all three of these validation methods together has allowed us to confirm the generation of both homozygous and heterozygous HS2ST1 functional knockouts using RNP delivery of CRISPR/Cas9. One particularly interesting finding is that the HS composition of HS2ST1$^{-/-}$ and HS2ST1$^{+/-}$ hiPSCs is similar to that of KO mice and patient 1, with a vast increase in UA-GlcNS(6S) disaccharides (between 31% and 36% across both cells lines in conditioned medium and cell lystaes).

Chapter 5 describes the phenotypic characterization of the HS2ST1$^{-/-}$ hiPSCs and begins to investigate their differentiation potential. Despite hiPSCs lacking what was thought to be an essential binding motif of FGF2, the cells were able to maintain their pluripotency and differentiate to early-stage endoderm and ectoderm lineages. It was found that FGF2 binding affinity was not altered, yet the downstream responses were reduced in *Hs2st1*$^{-/-}$ mESCs when compared to WT (Merry *et al.*, 2001). Additionally, a study into hiPSCs deficient in HS found that cells were still able to signal via FGF2 (Jamie Thompson, personal communication).

A key finding of this doctoral thesis is that, despite clear and significant differences in their HS structure, hiPSCs lacking HS2ST1 activity are still able to differentiate to DE similar to WT. Although some differences were seen in the degree of changes in gene expression during the differentiation protocols, it is not possible to draw firm conclusion as, unfortunately, they were only able to be carried out once. Furthermore, in HS2ST1$^{+/-}$ hiPSCs, levels of SOX17 appear increased when compared to WT and in HS2ST1$^{-/-}$ hiPSCs, SOX17 and FOXA2 levels appear reduced. However, these results did not correlate with immunostaining data. For that reason, it would be difficult to draw concrete conclusions regarding the impact of loss of HS2ST1 activity on specific differentiation processes and further repeats are required. However, it has previously been concluded that HS-deficient hiPSCs differentiate to DE faster than WT (Jamie Thompson, personal communication).

In addition to DE, Chapter 5 also discusses testing a neural protocol, which was optimized for successful differentiation for 7 days. As no GAG-free neural differentiation methods were available at the time of this study, this can be considered a success as several other methods have been tested, but failed to be reproducible (Jamie Thompson, personal communication). It is important to note that the protocol presented in this thesis needs to be carried out additional times in order to assess its reproducibility. To generate a longer neural protocol, further optimization would be required.

6.2 Future Work

Figure 46 suggests future experiments that could be planned as a follow up to the HS2ST1 study carried out during this doctoral thesis. In addition, this figure contains an overview of the experimental work that was carried out.

As part of this doctoral thesis, the HS composition of hiPSCs lacking HS2ST1 was investigated. Notable changes in the proportion of different HS disaccharides were seen. It would be interesting to determine whether altered HS patterning has any effect on the amount of other sulphated GAGs, or their composition. In hiPSCs lacking HS, CS/DS sulphation patterning was altered and the amount of KS present in the hiPSCs was vastly increased. The sulphation patterning of the KS was not able to be determined, but it was hypothesized that the composition of KS would also be altered (Jamie Thompson, personal communication). CD/DS composition could be investigated using disaccharide digestion, AMAC labelling and RP-HPLC separation. HS2ST1$^{+/-}$ and HS2ST1$^{-/-}$ hiPSCs could be metabolically radiolabelled which would provide information on chain length and domain organisation. Furthermore, this method would give an insight into the representative populations of other sulphated GAGs and so it could be determined whether CD/DS or KS production is upregulated when compared with WT. Furthermore, 3D-Orbi-SIMS (introduced in Chapter 1) could also be used which combines the spatial resolution of ToF-SIMS and the mass resolution of the Orbitrap which would not only allow investigation into how the GAGs have changed in

composition, but how they have also changed in their localisation within and around the cells. Additionally, GAG discriminatory peaks would allow us to understand whether individual GAG populations are changing in KO cell lines.

Although the gene expression levels of *FGFRs* were determined, this does not provide any information the FGFR protein levels. It is possible that despite the decrease in *FGFR3* mRNA, the protein levels are unchanged. It is important to note that FGFRs are known to generate structural diversity by alternative splicing; FGFR3 has 2 known variants (Wuechner *et al.*, 1996). This means that qRT-PCR may not pick up all of the splice variants and mRNA levels could appear reduced when, in fact, an alternative splice variant is upregulated. Additionally, an FGF2 signalling assay can be performed in which hiPSCs are treated with varying concentrations of FGF2 and the pERK1/2 quantified by western blot. E8 medium contains high levels of FGF2 so KO hiPSCs would be receiving an excess of FGF2 constantly. This excess could result in an FGF2 response despite reduced signalling. Reducing FGF2 concentrations in the medium would provide a better understanding of the FGF2 signalling capacity of the cells. When this assay was carried out using *Hs2st1*$^{-/-}$ mESCs, it was found that the KO mESCs elicited a reduced response to FGF2, yet no differences in binding affinity were observed (Merry *et al.*, 2001). This is a surprising finding, considering the HS 2-O-sulphation modification is thought to confer FGF2 binding, but indicates 2-O-sulphation may be required to elicit a response. However, it is hypothesized that reduced FGF2 response would be seen in HS2ST1$^{-/-}$ hiPSCs.

HS2ST1$^{-/-}$ and HS2ST1$^{+/-}$ hiPSCs are able to differentiate to early endoderm and ectoderm lineages, but a mesoderm differentiation would allow conclusions to be drawn regarding tri-lineage differentiation capabilities. In order to gain a fuller understanding, more detailed investigations into the phenotypic characterization could be carried out. Only immunostaining, flow cytometry and qRT-PCR have been performed as part of this doctoral thesis, but characterisation of signalling pathways or cell function could be performed to determine any differences between cell lines. Functional assays might include electrophysical studies of neural networks or contraction measurements in cardiomyocytes

(Ahlenius *et al.*, 2009; Jian *et al.*, 2014; Chun *et al.*, 2021). Furthermore, differentiating the cells for longer would also be of interest as it could be that, over short periods of time, differences between KO and WT hiPSCs are not clear. Optimizing a neural protocol would be a priority as these differentiations are often lengthy and could provide a good understanding of the neurological aspects of HS2ST1-associated disease. Furthermore, bone and kidney differentiations would allow characterisation of specific pathways or mechanisms that are altered in the absence of HS2ST1 as these are the main organs affected in mice models, and human patients. However, to date, no GAG-free differentiations for these tissues exist and published methods often require Matrigel® or heparin additions. Kidney organoids would be of interest in HS2ST1$^{-/-}$ hiPSCs as the abnormalities associated with kidney formation arise at an early stage of development and not in more mature podocyte cells, for example. However, the conditions in which these are cultured would need to be rigorously tested to ensure no exogenous GAGs are added into the system. Once the hiPSCs have been terminally differentiated, mRNA and protein changes can be determined, as has been shown in this study. Furthermore, 3D-Orbi-SIMS could be carried out to assess how the matrix changes in development. GAG disaccharide analysis would also be useful to gain an understanding of how GAGs change during differentiation/development and whether these changes are universal or tissue specific.

Transcriptomics and proteomics are becoming increasingly popular techniques to look at gene networks and protein functions of different hiPSC lines. Transcriptomic technologies, such as RNA-seq, would allow a greater insight into gene structure, expression and product function (Dong and Chen, 2013). In this project, RNA-seq would provide information on global transcriptional changes in the HS2ST1 altered cells lines which could then be validated by qPCR. As little information is known about the regulation of GAGs and their biosynthetic enzymes, RNA-seq of KO hiPSCs could provide insight into GAG-related genes which have not been picked up to date. Such discoveries could then be investigated further by exploring the effect of the loss of other, seemingly related genes in GAG biosynthesis. Proteomic analysis provides further detail on the interaction, localization and post translational modifications (such as phosphorylation) of proteins (Reinders and Sickmann, 2005;

Aslam *et al.*, 2017). As protein phosphorylation plays a large role in hiPSCs differentiation and pluripotency, both proteomic and phosphoproteomic analysis would be of great interest to determine whether the loss of an 'essential' HS modification has any effects on signalling pathways. Furthermore, as GAGs are attached to a protein core, proteomics would provide information on how core proteins are differentially expressed when GAGs have alternative sulphation patterns. Proteomics would allow an investigation into how GAG structure affects the PG core protein as well as its localization within the cells.

Another particularly interesting experiment would be the generation of hiPSC lines from patient-derived material. The advantage of this is that the genetic background of the patients would be maintained. This means that if there are other factors, undetected by a full genome scan, they would still have an effect on the cells. Once generated, the pluripotency of the hiPSCs would need to be assessed and the cells characterized in the same way as described in this doctoral thesis (including potential future experiments outlines in this chapter). This would provide a wealth of knowledge regarding GAG changes and alterations in specific developmental or signalling pathways in HS2ST1-associated disease. In addition, creating isogenic controls of patient derived hiPSCs by using CRISPR/Cas9 and HDR would also be important. These isogenic control lines would also need to be characterized in a similar way to the patient derived hiPSCs and CRISPR/Cas9 $HS2ST1^{-/-}$ and $HS2ST1^{+/-}$ hiPSCs.

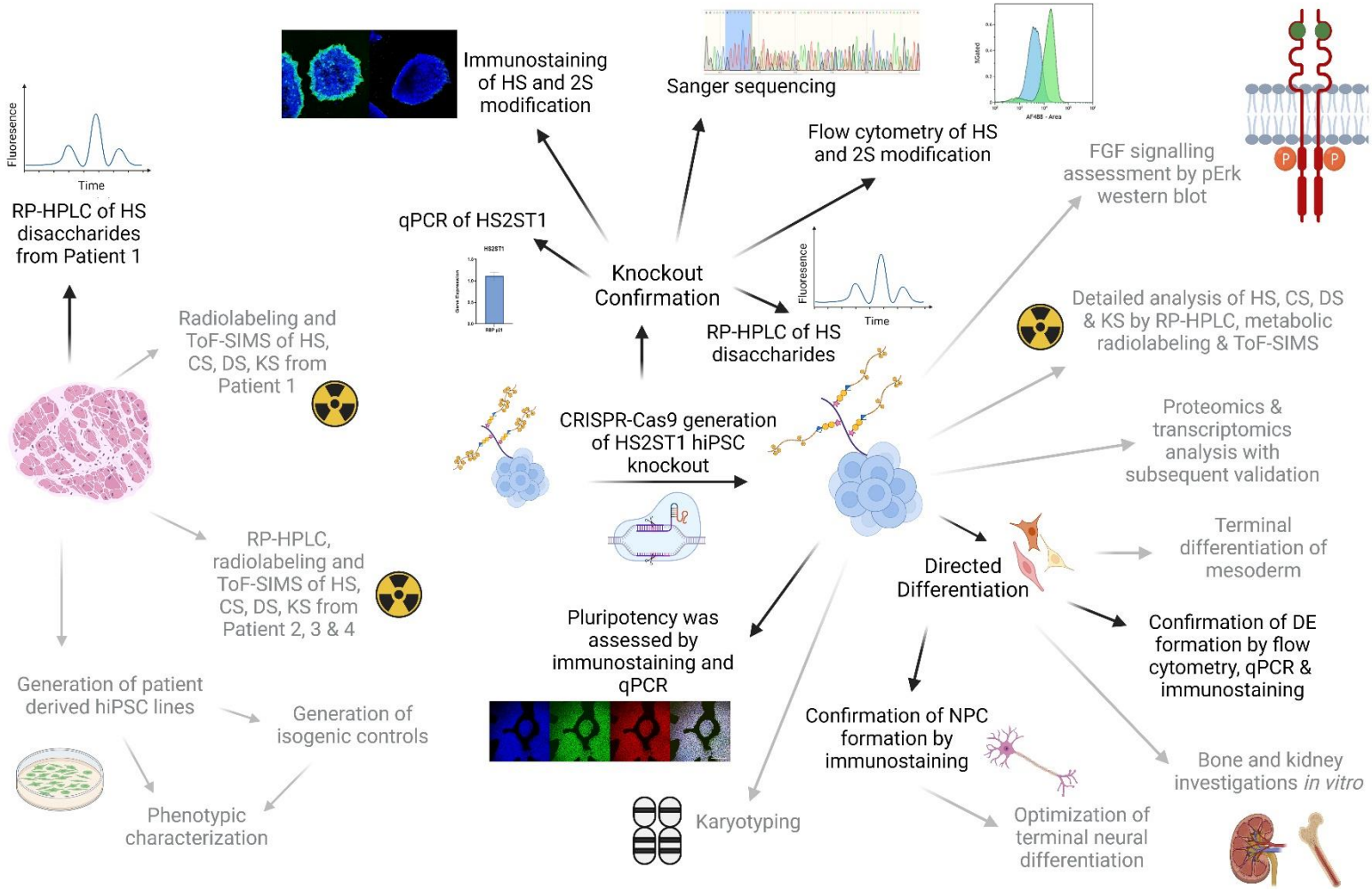


Figure 46 - Summary of Work and Future Work. Summary of the work carried out in this doctoral thesis shown in black and suggested future experiments shown in grey.

References

Ahlenius, H. *et al.* (2009) 'Neural stem and progenitor cells retain their potential for proliferation and differentiation into functional neurons despite lower number in aged brain', *Journal of Neuroscience*, 29(14), pp. 4408–4419.

Aikawa, J. I. *et al.* (2001) 'Multiple isozymes of heparan sulfate/heparin GlcNAc N-deacetylase/GlcN N-sulfotransferase. Structure and activity of the fourth member, NDST4', *Journal of Biological Chemistry*, 276(8), pp. 5876–5882.

Anower-E-Khuda, M. F. *et al.* (2013) 'Heparan sulfate 6-O-sulfotransferase isoform-dependent regulatory effects of heparin on the activities of various proteases in mast cells and the biosynthesis of 6-O-sulfated heparin', *Journal of Biological Chemistry*, 288(6), pp. 3705–3717.

Apáti, Á., Berecz, T. and Sarkadi, B. (2016) 'Calcium signaling in human pluripotent stem cells', *Cell Calcium*, 59(2–3), pp. 117–123.

Archer-Hartmann, S. *et al.* (2023) 'Isolation and Compositional Analysis of Glycosaminoglycans', *Methods in Molecular Biology*, 2597, pp. 177–186.

Arfi, A. *et al.* (2011) 'Neuroinflammatory and oxidative stress phenomena in MPS IIIA mouse model: The positive effect of long-term aspirin treatment', *Molecular Genetics and Metabolism*, 103(1), pp. 18–25.

Armstrong, L. *et al.* (2017) 'A girl with developmental delay, ataxia, cranial nerve palsies, severe respiratory problems in infancy—Expanding NDST1 syndrome', *American Journal of Medical Genetics, Part A*, 173(3), pp. 712–715.

Aslam, B. *et al.* (2017) 'Proteomics: Technologies and their applications', *Journal of Chromatographic Science*, 55(2), pp. 182–196.

Audouard, E. *et al.* (2021) 'Complete Correction of Brain and Spinal Cord Pathology in Metachromatic Leukodystrophy Mice', *Frontiers in Molecular Neuroscience*, 14.

Bai, X. and Esko, J. D. (1996) 'An animal cell mutant defective in heparan sulfate hexuronic acid 2-O- sulfation', *Journal of Biological Chemistry*, 271(30), pp. 17711–17717.

Beck, M. (2018) 'Treatment strategies for lysosomal storage disorders', *Developmental Medicine and Child Neurology*, 60(1), pp. 13–18.

Bellaiche, Y., The, I. and Perrimon, N. (1998) 'Tout-velu is a Drosophila homologue of the putative tumour suppressor EXT-1 and is needed for Hh diffusion', *Nature*, 394(6688), pp. 85–88.

Beltrami, G. *et al.* (2016) 'Hereditary Multiple Exostoses: a review of clinical appearance and metabolic pattern', *Clinical Cases in Mineral and Bone Metabolism*, 13(2), pp. 110–118.

Benető, N. *et al.* (2020) 'Sanfilippo syndrome: Molecular basis, disease models and therapeutic approaches', *International Journal of Molecular Sciences*, 21(21), pp. 1–20.

Bink, R. J. *et al.* (2003) 'Heparan Sulfate 6-O-Sulfotransferase Is Essential for Muscle Development in Zebrafish', *The Journal of Biological Chemistry*, 278(33), pp. 31118–31127..

Boado, R. J. *et al.* (2014) 'Insulin receptor antibody-sulfamidase fusion protein penetrates the primate blood-brain barrier and reduces glycosoaminoglycans in sanfilippo type a cells', *Molecular Pharmaceutics*, 11(8), pp. 2928–2934.

Bond, A. M., Bhalala, O. G. and Kessler, J. A. (2012) 'The dynamic role of bone morphogenetic proteins in neural stem cell fate and maturation', *Developmental Neurobiology*, 72(7), pp. 1068–1084.

Bork, K. *et al.* (2021) 'Novel hereditary angioedema linked with a heparan sulfate 3-O-sulfotransferase 6 gene mutation', *Journal of Allergy and Clinical Immunology*, 148(4), pp. 1041–1048.

Bovée, J. V. M. G. (2008) 'Multiple osteochondromas', *Orphanet Journal of Rare Diseases*, 3(1).

Bullock, S. L. *et al.* (1998) 'Renal agenesis in mice homozygous for a gene trap mutation in the gene encoding heparan sulfate 2-sulfotransferase', *Genes and Development*, 12(12), pp. 1894–1906.

Bülow, H. E. and Hobert, O. (2004) *Differential Sulfations and Epimerization Define Heparan Sulfate Specificity in Nervous System Development*, *Neuron*. Available at: http://www.neuron.org/cgi/content/full/41/5/ (Accessed: 25 June 2020).

Busse, M. *et al.* (2007) 'Contribution of EXT1, EXT2, and EXTL3 to heparan sulfate chain elongation', *Journal of Biological Chemistry*, 282(45), pp. 32802–32810.

Caetano-Lopes, J., Canhão, H. and Fonseca, J. E. (2007) 'Osteoblasts and bone formation.', *Acta reumatológica portuguesa*, 32(2), pp. 103–110.

Caplan, A. I. (1991) 'Mesenchymal stem cells', *Journal of Orthopaedic Research*, 9(5), pp. 641–650.

Carey, D. J. (1997) 'Syndecans : multifunctional cell-surface co-receptors', *Biochem. J*, 327, pp. 1–16.

de Carvalho, T. G. *et al.* (2018) 'CRISPR-Cas9-mediated gene editing in human MPS I fibroblasts', *Gene*, 678, pp. 33–37.

De Cat, B. and David, G. (2001) 'Developmental roles of the glypicans', *Seminars in Cell and Developmental Biology*, 12(2), pp. 117–125.

Chen, G. *et al.* (2011) 'Chemically defined conditions for human iPSC derivation and culture', *Nature Methods*, 8(5), pp. 424–429.

Chen, J. *et al.* (2003) 'Biosynthesis of 3-O-sulfated heparan sulfate: unique substrate specificity of heparan sulfate 3-O-sulfotransferase isoform 5', *Glycobiology*, 13(11), pp. 785–794.

Chen, Y.-H. *et al.* (2018) 'The GAGOme: a cell-based library of displayed glycosaminoglycans', *Nature Methods*, 15, pp. 881–888.

Christensen, C. and Choy, F. (2017) 'A Prospective Treatment Option for Lysosomal Storage Diseases: CRISPR/Cas9 Gene Editing Technology for Mutation Correction in Induced Pluripotent Stem Cells', *Diseases*.

Diseases, 5(1).

Chun, C. *et al.* (2021) 'Astrocyte-derived extracellular vesicles enhance the survival and electrophysiological function of human cortical neurons in vitro', *Biomaterials*, 271(120700).

Clarke, D. *et al.* (2018) 'Genetically Corrected iPSC-Derived Neural Stem Cell Grafts Deliver Enzyme Replacement to Affect CNS Disease in Sanfilippo B Mice', *Molecular Therapy - Methods and Clinical Development*, 10, pp. 113–127.

Cleary, M. A. and Wraith, J. E. (1993) 'Management of mucopolysaccharidosis type III', *Archives of Disease in Childhood*, 69(3), pp. 403–406.

Concolino, D., Deodato, F. and Parini, R. (2018) 'Enzyme replacement therapy: Efficacy and limitations', *Italian Journal of Pediatrics*, 44(2).

Cortez, C. and Patrick, M. (2015) *CRISPR 101: Homology Directed Repair*, *Addgene's Blog*. Available at: http://blog.addgene.org/crispr-101-homology-directed-repair (Accessed: 25 January 2022).

Crawford, B. E. *et al.* (2001) 'Cloning, Golgi Localization, and Enzyme Activity of the Full-length Heparin/Heparan Sulfate-Glucuronic Acid C5-epimerase', *Journal of Biological Chemistry*, 276(24), pp. 21538–21543.

Dagälv, A. *et al.* (2022) 'Heparan Sulfate Structure: Methods to Study N-Sulfation and NDST Action', in Balahurunathan, K. et al. (eds) *Methods in Molecular Biology*. 2nd edn. Humana Press, pp. 139–150.

Dalton, S. (2013) 'Signaling networks in human pluripotent stem cells', *Current Opinion in Cell Biology*, 25(2), pp. 241–246.

David, G. *et al.* (1992) 'Developmental changes in heparan sulfate expression: In situ detection with mAbs', *Journal of Cell Biology*, 119(4), pp. 961–975.

Deakin, J. A. and Lyon, M. (2008) 'A simplified and sensitive fluorescent method for disaccharide analysis of both heparan sulfate and chondroitin/dermatan sulfates from biological samples', *Glycobiology*, 18(6), pp. 483–491.

Deng, C. *et al.* (1996) 'Fibroblast growth factor receptor 3 is a negative regulator of bone growth', *Cell*, 84(6), pp. 911–921.

Denys, A. and Allain, F. (2019) 'The emerging roles of heparan sulfate 3-O-sulfotransferases in cancer', *Frontiers in Oncology*, 9(507).

Dhoot, G. K. *et al.* (2001) 'Regulation of Wnt signaling and embryo patterning by an extracellular sulfatase', *Science*, 293(5535), pp. 1663–1666.

Dong, Z. C. and Chen, Y. (2013) 'Transcriptomics: Advances and approaches', *Science China Life Sciences*, 56(10), pp. 960–967.

Ebert, A. D. *et al.* (2009) 'Induced pluripotent stem cells from a spinal muscular atrophy patient', *Nature*, 457(7227), pp. 277–280.

Esko, J. D., Kimata, K. and Lindahl, U. (2009) 'Proteoglycans and Sulfated Glycosaminoglycans', in Varki, A. et al. (eds) *Essentials of Glycobiology*. 2nd edn. Cold Spring Harbor Laboratory Press. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/20301236 (Accessed: 28 May 2020).

Esko, J. D. and Lindahl, U. (2001) 'Molecular diversity of heparan sulfate', *Journal of Clinical Investigation*, 108(2), pp. 169–173.

Evans, M. J. and Kaufman, M. H. (1981) 'Establishment in culture of pluripotential cells from mouse embryos', *Nature*, 292(5819), pp. 154–156.

Faa, G. *et al.* (2012) 'Morphogenesis and molecular mechanisms involved in human kidney development', *Journal of Cellular Physiology*, 227(3), pp. 1257–1268.

Fan, G. *et al.* (2000) 'Targeted disruption of NDST-1 gene leads to pulmonary hypoplasia and neonatal respiratory distress in mice', *FEBS Letters*, 467(1), pp. 7–11.

Fatehullah, A., Tan, S. H. and Barker, N. (2016) 'Organoids as an in vitro model of human development and disease', *Nature Cell Biology*, 18(3), pp. 246–254.

Fawcett, J. W. *et al.* (2022) 'The extracellular matrix and perineuronal nets in memory', *Molecular Psychiatry 2022*, 27(8), pp. 3192–3203.

Fecarotta, S., Gasperini, S. and Parenti, G. (2018) 'New treatments for the mucopolysaccharidoses: From pathophysiology to therapy 11 Medical and Health Sciences 1103 Clinical Sciences', *Italian Journal of Pediatrics*, 44(2)

Fedele, A. O. (2015) 'Sanfilippo syndrome: Causes, consequences, and treatments', *Application of Clinical Genetics*, 8, pp. 269–281.

Federico, A. and de Visser, M. (2021) 'New disease modifying therapies for two genetic childhood-onset neurometabolic disorders (metachromatic leucodystrophy and adrenoleucodystrophy)', *Neurological Sciences*, 42(7), pp. 2603–2606.

Feng, R. and Wen, J. (2015) 'Overview of the roles of Sox2 in stem cell and development', *Biological Chemistry*, 396(8), pp. 883–891.

Filmus, J., Capurro, M. and Rast, J. (2009a) 'Glypicans', *Genome Biology*, 9(5), pp. 5–10.

Filmus, J., Capurro, M. and Rast, J. (2009b) 'Glypicans', *Genome Biology*., 9(5), pp. 224.

Forsberg, E. *et al.* (1999) Abnormal mast cells in mice deficient in a heparin-synthesizing enzyme, *Nature,* 400(6746), pp.773-776.

Forsberg, E. and Kjellén, L. (2001) 'Heparan sulfate: lessons from knockout mice', *Journal of Clinical Investigation*, 108(2), pp. 175–180.

Fraldi, A. *et al.* (2007) 'Functional correction of CNS lesions in an MPS-IIIA mouse model by intracerebral AAV-mediated delivery of sulfamidase and SUMF1 genes', *Human Molecular Genetics*, 16(22), pp. 2693–2702.

Fransson, L. Å. (2003) 'Glypicans', *International Journal of Biochemistry and Cell Biology*, 35(2), pp. 125–129..

Frantz, C., Stewart, K. M. and Weaver, V. M. (2010) 'The extracellular

matrix at a glance', *Journal of Cell Science*, 123(24), pp. 4195–4200.

Freeze, H. H., Kinoshita, T. and Schnaar, R. L. (2017) 'Genetic Disorders of Glycan Degradation', in Varki, A. et al. (eds) *Essentials of Glycobiology*. 3rd edn. Cold Spring Harbor Laboratory Press.

Frega, M. *et al.* (2017) 'Rapid neuronal differentiation of induced pluripotent stem cells for measuring network activity on micro-electrode arrays', *Journal of Visualized Experiments*, 2017(119).

Frese, M. A. *et al.* (2009) 'Characterization of the human sulfatase Sulf1 and its high affinity heparin/heparan sulfate interaction domain', *Journal of Biological Chemistry*, 284(41), pp. 28033–28044.

Gaffke, L. *et al.* (2018) 'How close are we to therapies for Sanfilippo disease?', *Metabolic Brain Disease*, 33(1).

Galliher, P. M. *et al.* (1981) 'Heparinase production by Flavobacterium heparinum.', *Applied and Environmental Microbiology*, 41(2), pp. 360.

George, N. and Geller, H. M. (2018) 'Extracellular matrix and traumatic brain injury', *Journal of Neuroscience Research*, 96(4), pp. 573–588.

Gerrard, L., Rodgers, L. and Cui, W. (2005) 'Differentiation of Human Embryonic Stem Cells to Neural Lineages in Adherent Culture by Blocking Bone Morphogenetic Protein Signaling', *Stem Cells*, 23(9), pp. 1234–1241.

Ginis, I. *et al.* (2004) 'Differences between human and mouse embryonic stem cells', *Developmental Biology*, 269(2), pp. 360–380.

Gomez-Ospina, N. (1993) 'Arylsulfatase A Deficiency', *GeneReviews®*. University of Washington, Seattle. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20301309 (Accessed: 19 March 2023).

Gondelaud, F. and Ricard-Blum, S. (2019) 'Structures and interactions of syndecans', *FEBS Journal*, 286(15), pp. 2994–3007.

Gostimskaya, I. (2022) 'CRISPR–Cas9: A History of Its Discovery and Ethical Considerations of Its Use in Genome Editing', *Biochemistry*, 87(8), pp. 777.

Götting, C., Kuhn, J. and Kleesiek, K. (2007) 'Human xylosyltransferases in health and disease', *Cellular and Molecular Life Sciences*, 64(12), pp. 1498–1517.

Grapin-Botton, A. and Constam, D. (2007) 'Evolution of the mechanisms and molecular control of endoderm formation', *Mechanisms of Development*, 124(4), pp. 253–278.

Grobe, K. *et al.* (2005) 'Cerebral hypoplasia and craniofacial defects in mice lacking heparan sulfate Ndst1 gene function', *Development*, 123(16), pp. 3777–3786.

Guimond, S. *et al.* (1993) 'Activating and inhibitory heparin sequences for FGF-2 (basic FGF). Distinct requirements for FGF-1, FGF-2, and FGF-4', *J Biol Chem*, 268(32), pp. 23906–23914.

Gulberti, S. *et al.* (2020) 'The role of heparan sulfate maturation in cancer: A focus on the 3O-sulfation and the enigmatic 3O-

sulfotransferases (HS3STs)', *Seminars in Cancer Biology*, 62, pp. 68–85.

Gupta, R. M. and Musunuru, K. (2014) 'Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9', *Journal of Clinical Investigation*, 124(10), pp. 4154–4161.

Habuchi, H. *et al.* (2000) 'The Occurrence of Three Isoforms of Heparan Sulfate 6-O-Sulfotransferase Having Different Specificities for Hexuronic Acid Adjacent to the Targeted N-Sulfoglucosamine', *The Journal of Biological Chemistry*, 275(4), pp. 2859–2868..

Habuchi, H. *et al.* (2007) 'Mice deficient in heparan sulfate 6-O-sulfotransferase-1 exhibit defective heparan sulfate biosynthesis, abnormal placentation, and late embryonic lethality', *Journal of Biological Chemistry*, 282(21), pp. 15578–15588..

Habuchi, H., Kobayashi, M. and Kimata, K. (1998) Molecular characterization and expression of heparan-sulfate 6- sulfotransferase. Complete cDNA cloning in human and partial cloning in Chinese hamster ovary cells, *Journal of Biological Chemistry,* 273(15), pp.9208-9213.

Häcker, U., Lin, X. and Perrimon, N. (1997) 'The Drosophila sugarless gene modulates Wingless signaling and encodes an enzyme involved in polysaccharide biosynthesis', *Development*, 124(18), pp. 3565–3573.

Hagner-McWhirter, A. (2000) 'Biosynthesis of heparin/heparan sulfate: kinetic studies of the glucuronyl C5-epimerase with N-sulfated derivatives of the Escherichia coli K5 capsular polysaccharide as substrates', *Glycobiology*, 10(2), pp. 159–171.

Hagner-McWhirter, Å. *et al.* (2004) 'Irreversible Glucuronyl C5-epimerization in the Biosynthesis of Heparan Sulfate', *Journal of Biological Chemistry*, 279(15), pp. 14631–14638.

HajMohammadi, S. *et al.* (2003) 'Normal levels of anticoagulantheparan sulfate are not essential for normal hemostasis', *The Journal of Clinical Investigation*, 111(7).

Halder, S. K., Sapkota, A. and Milner, R. (2022) 'The impact of genetic manipulation of laminin and integrins at the blood–brain barrier', *Fluids and Barriers of the CNS*, 19(1), pp. 1–12.

Hascall, V. C., Weigel, P. H. and Toole, B. P. (2016) 'Hyaluronan', *Encyclopedia of Cell Biology*, 1, pp. 279–287.

Hemmati-Brivanlou, A. and Meltont, D. (1997) 'Vertebrate embryonic cells will become nerve cells unless told otherwise', *Cell*, 88(1), pp. 13–17.

Hirabayashi, Y. *et al.* (2004) 'The Wnt/β-catenin pathway directs neuronal differentation of cortical neural precursor cells', *Development*, 131(12), pp. 2791–2801.

Holley, R. J., Pickford, C. E., *et al.* (2011) 'Influencing hematopoietic differentiation of mouse embryonic stem cells using soluble heparin and heparan sulfate saccharides', *Journal of Biological Chemistry*, 286(8), pp. 6241–6252.

Holley, R. J., Deligny, A., *et al.* (2011) 'Mucopolysaccharidosis type I, unique structure of accumulated heparan sulfate and increased N-

sulfotransferase activity in mice lacking α-L-iduronidase', *Journal of Biological Chemistry*, 286(43), pp. 37515–37524.

Holmborn, K. *et al.* (2004) 'Heparan sulfate synthesized by mouse embryonic stem cells deficient in NDST1 and NDST2 is 6-O-sulfated but contains no N-sulfate groups', *Journal of Biological Chemistry*, 279(41), pp. 42355–42358.

Holst, C. R. *et al.* (2007) 'Secreted sulfatases Sulf1 and Sulf2 have overlapping yet essential roles in mouse neonatal survival', *PLoS ONE*, 2(6).

Hook, A. L. *et al.* (2021) 'High sensitivity analysis of nanogram quantities of glycosaminoglycans using ToF-SIMS', *Communications Chemistry*, 4(1).

Hoshiba, T. *et al.* (2016) 'Decellularized extracellular matrix as an in vitro model to study the comprehensive roles of the ECM in stem cell differentiation', *Stem Cells International*, 2016.

Hou, J. *et al.* (2018) 'Retaining mTeSR1 medium during hepatic differentiation facilitates hepatocyte-like cell survival by decreasing apoptosis', *Cellular Physiology and Biochemistry*, 51(4), pp. 1533–1543.

Hu, B. Y. *et al.* (2010) 'Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency', *Proceedings of the National Academy of Sciences of the United States of America*, 107(9), pp. 4335–4340.

Hu, W. *et al.* (2016) 'Suppression of Nestin reveals a critical role for p38-EGFR pathway in neural progenitor cell proliferation', *Oncotarget*. Impact Journals, 7(52), pp. 87052–87063.

Hynes, R. O. (2009) 'The extracellular matrix: Not just pretty fibrils', *Science*, 326(5957), pp. 1216–1219.

Iozzo, R. V and Schaefer, L. (2015) 'Proteoglycan form and function: A comprehensive nomenclature of proteoglycans', *Matrix Biol*, 42, pp. 11–55.

Ishiguro, K. *et al.* (2000) 'Complete antithrombin deficiency in mice results in embryonic lethality', *Journal of Clinical Investigation*, 106(7), pp. 873–878.

Jacob, A. *et al.* (2019) 'Derivation of self-renewing lung alveolar epithelial type II cells from human pluripotent stem cells HHS Public Access', *Nat Protoc*, 14(12), pp. 3303–3332.

Jacobsson, I. *et al.* (1979) Identification of N-sulphated disaccharide units in heparin-like polysaccharides, *Biochemical Journal, 179(1), pp.77-87*.

Jao, T. M. *et al.* (2016) 'Alteration of colonic epithelial cell differentiation in mice deficient for glucosaminyl N-deacetylase/N-sulfotransferase 4', *Oncotarget*, 7(51), pp. 84938–84950.

Jemth, P. *et al.* (2003) 'Oligosaccharide Library-based Assessment of Heparan Sulfate 6-O-Sulfotransferase Substrate Specificity', *Journal of Biological Chemistry*, 278(27), pp. 24371–24376.

Jenniskens, G. J. *et al.* (2000) 'Heparan sulfate heterogeneity in skeletal muscle basal lamina: Demonstration by phage display-derived antibodies', *Journal of Neuroscience*, 20(11), pp. 4099–4111.

Jenniskens, G. J. *et al.* (2002) 'Spatiotemporal distribution of heparan sulfate epitopes during myogenesis and synaptogenesis: A study in developing mouse intercostal muscle', *Developmental Dynamics*, 225(1), pp. 70–79.

Jian, Z. *et al.* (2014) 'Mechanochemotransduction during cardiomyocyte contraction is mediated by localized nitric oxide signaling', *Science Signaling*, 7(317).

Jiang, F. and Doudna, J. A. (2017) 'CRISPR-Cas9 Structures and Mechanisms', *Annual Review of Biophysics*, pp. 505–529.

Jinek, M. *et al.* (2012) 'A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity', *Science*, 337(6096), pp. 816–821.

Johnson, C. E. *et al.* (2007) 'Essential Alterations of Heparan Sulfate During the Differentiation of Embryonic Stem Cells to Sox1-Enhanced Green Fluorescent Protein-Expressing Neural Progenitor Cells', *Stem Cells*, 25(8), pp. 1913–1923.

Justice, M. J. and Dhillon, P. (2016) 'Using the mouse to model human disease: Increasing validity and reproducibility', *DMM Disease Models and Mechanisms*, 9(2), pp. 101–103.

Kalus, I. *et al.* (2009) 'Differential involvement of the extracellular 6-O-endosulfatases Sulf1 and Sulf2 in brain development and neuronal and behavioural plasticity', *Journal of Cellular and Molecular Medicine*, 13(11–12), pp. 4505–4521.

Kamimura, K. *et al.* (2001) 'Drosophila Heparan Sulfate 6-O-Sulfotransferase (dHS6ST) Gene', *The Journal of Biological Chemistry*, 276(20), pp. 17014–17021.

Karlsson, R. *et al.* (2021) 'Dissecting structure-function of 3-O-sulfated heparin and engineered heparan sulfates', *Science Advances*, 7(52), pp. 6026.

Katta, K. *et al.* (2015) 'Reduced expression of EXTL2, a member of the exostosin (EXT) family of glycosyltransferases, in human embryonic kidney 293 cells results in longer heparan sulfate chains', *Journal of Biological Chemistry*, 290(21), pp. 13168–13177.

Khan, S. A. *et al.* (2008) 'Endogenous heparan sulfate and heparin modulate bone morphogenetic protein-4 signaling and activity', *American Journal of Physiology*, 294(6), pp. 1387–1397.

Kiecker, C. and Niehrs, C. (2001) 'A morphogen gradient of Wnt/β-catenin signalling regulates anteroposterior neural patterning in Xenopus', *Development*, 128(21), pp. 4189–4201.

Kim, D. S. *et al.* (2010) 'Robust enhancement of neural differentiation from human ES and iPS cells regardless of their innate difference in differentiation propensity', *Stem Cell Reviews and Reports*, 6(2), pp. 270–281.

Kim, E. J., Kang, K. H. and Ju, J. H. (2017) 'Crispr-cas9: A promising tool for gene editing on induced pluripotent stem cells', *Korean Journal of Internal Medicine*, 32(1), pp. 42–61.

Kjellén, L. (2003) 'Glucosaminyl N-deacetylase/N-sulphotransferases in heparan sulphate biosynthesis and biology', *Biochemical Society Transactions*, 31(2), pp. 340–342.

Kobayashi, M. *et al.* (1996) 'Purification and characterization of heparan sulfate 2-sulfotransferase from cultured Chinese hamster ovary cells', *Journal of Biological Chemistry*, 271(13), pp. 7645–7653.

Kraushaar, D. C., Yamaguchi, Y. and Wang, L. (2010) 'Heparan sulfate is required for embryonic stem cells to exit from self-renewal', *Journal of Biological Chemistry*, 285(8), pp. 5907–5916.

Kresse, H. and Glössl, J. (1987) 'Glycosaminoglycan Degradation', in Meister, A. (ed.) *Advances in Enzymology and Related Areas of Molecular Biology*, pp. 217–311.

Kuo, W. J., Digman, M. A. and Lander, A. D. (2010) 'Heparan sulfate acts as a bone morphogenetic protein coreceptor by facilitating ligand-induced receptor hetero-oligomerization', *Molecular Biology of the Cell*, 21(22), pp. 4028–4041.

Van Kuppevelt, T. H. *et al.* (1998) 'Generation and application of type-specific anti-heparan sulfate antibodies using phage display technology: Further evidence for heparan sulfate heterogeneity in the kidney', *Journal of Biological Chemistry*, 273(21), pp. 12960–12966.

Kurup, S. *et al.* (2007) 'Characterization of anti-heparan sulfate phage display antibodies AO4B08 and HS4E4', *Journal of Biological Chemistry*, 282(29), pp. 21032–21042.

Kusche-Gullberg, M. *et al.* (1998) Identification and expression in mouse of two heparan sulfate glucosaminyl N-deacetylase/N-sulfotransferase genes, *Journal of Biological Chemistry,* 273(19), pp. 11902-11907.

LaFountaine, J. S., Fathe, K. and Smyth, H. D. C. (2015) 'Delivery and therapeutic applications of gene editing technologies ZFNs, TALENs, and CRISPR/Cas9', *International Journal of Pharmaceutics*, 494(1), pp. 180–194.

Lalonde, S. *et al.* (2017) 'Frameshift indels introduced by genome editing can lead to in-frame exon skipping', *PLoS ONE*, 12(6).

Lamanna, W. C. *et al.* (2008) 'Sulf loss influences N-, 2-O-, and 6-O-sulfation of multiple heparan sulfate proteoglycans and modulates fibroblast growth factor signaling', *Journal of Biological Chemistry*, 283(41), pp. 27724–27735.

Lamb, T. M. *et al.* (1993) 'Neural induction by the secreted polypeptide noggin', *Science*, 262(5134), pp. 713–718.

Langford-Smith, A. *et al.* (2012) 'Hematopoietic stem cell and gene therapy corrects primary neuropathology and behavior in mucopolysaccharidosis IIIA mice', *Molecular Therapy*, 20(8), pp. 1610–1621.

Laurent, L. C. *et al.* (2011) 'Dynamic changes in the copy number of

pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture', *Cell Stem Cell*, 8(1), pp. 106–118.

Lawrence, R. *et al.* (2007) 'The principal neuronal gD-type 3-O-sulfotransferases and their products in central and peripheral nervous system tissues', *Matrix Biology*, 26(6), pp. 442–455.

Lehmann, R. J. *et al.* (2021) 'Impaired neural differentiation of MPS IIIA patient induced pluripotent stem cell-derived neural progenitor cells', *Molecular Genetics and Metabolism Reports*, 29.

Leisico, F. *et al.* (2022) 'Structure of the human heparan sulfate polymerase complex EXT1-EXT2', *Nature Communications*, 13(1).

Li, J. P. *et al.* (2003) 'Targeted disruption of a murine glucuronyl C5-epimerase gene results in heparan sulfate lacking L-iduronic acid and in neonatal lethality', *Journal of Biological Chemistry*, 278(31), pp. 28363–28366.

Li, X. *et al.* (2018) 'A novel mutation of SGSH and clinical features analysis of mucopolysaccharidosis type IIIA', *Medicine (United States)*, 97(52).

Li, Y. *et al.* (2020) 'An optimized method for neuronal differentiation of embryonic stem cells in vitro', *Journal of Neuroscience Methods*, 330(108486).

Limtiaco, J. F. K., Jones, C. J. and Larive, C. K. (2009) 'Characterization of heparin impurities with HPLC-NMR using weak anion exchange chromatography', *Analytical Chemistry*, 81(24), pp. 10116–10123.

Lin, X. *et al.* (1999) 'Heparan sulfate proteoglycans are essential for FGF receptor signaling during Drosophila embryonic development', *Development*, 126(17), pp. 3715–3723.

Lin, X. *et al.* (2000) 'Disruption of gastrulation and heparan sulfate biosynthesis in EXT1-deficient mice', *Developmental Biology*, 224(2), pp. 299–311.

Lindahl, U. *et al.* (1980) *Evidence for a 3-O-sulfated D-glucosamine residue in the antithrombin-binding sequence of heparin*, *Biochemistry*, 77(11), pp.6551-6555.

Linker, A. and Hovingh, P. (1968) *The enzymatic degradation of heparitin sulfate*, *Biochimica et Biophysica Acta (BBA) - General Subjects,* 1473(1), pp.1-3.

Liu, A. and Niswander, L. A. (2005) 'Bone morphogenetic protein signalling and vertebrate nervous system development', *Nature Reviews Neuroscience*, 6(12), pp. 945–954.

Liu, C. *et al.* (2017) 'Delivery strategies of the CRISPR-Cas9 gene-editing system for therapeutic applications', *Journal of Controlled Release*, 266, pp. 17–26.

Liu, J. *et al.* (1999) 'Expression of heparan sulfate D-glucosaminyl 3-O-sulfotransferase isoforms reveals novel substrate specificities', *Journal of Biological Chemistry*, 274(8), pp. 5185–5192.

Maccarana, M. *et al.* (1996) Domain structure of heparan sulfates from

bovine organs, *Journal of Biological Chemistry, 271(30), pp.17804-17810*.

Maccarana, M., Casu, B. and Lindahl, U. (1993) 'Minimal sequence in heparin/heparan sulfate required for binding of basic fibroblast growth factor', *Journal of Biological Chemistry*, 268(32), pp. 23898–23905.

El Masri, R. *et al.* (2022) 'Extracellular endosulfatase Sulf-2 harbors a chondroitin/dermatan sulfate chain that modulates its enzyme activity', *Cell Reports*, 38(11).

de Mattos Pimenta Vidal, C. *et al.* (2017) 'Role of proteoglycans on the biochemical and biomechanical properties of dentin organic matrix', *Archives of Oral Biology*, 82, pp. 203–208.

Mayfosh, A. J., Nguyen, T. K. and Hulett, M. D. (2021) 'The Heparanase Regulatory Network in Health and Disease', *International journal of molecular sciences*, 22(20).

McCormick, C. *et al.* (2000) The putative tumor suppressors EXT1 and EXT2 form a stable complex that accumulates in the Golgi apparatus and catalyzes the synthesis of heparan sulfate, *Proceedings of the National Academy of Sciences of the United States of America,* 97(2), pp.668-673.

McGlynn, R., Dobrenis, K. and Walkley, S. U. (2004) 'Differential subcellular localization of cholesterol, gangliosides, and glycosaminoglycans in murine models of mucopolysaccharide storage disorders', *Journal of Comparative Neurology*, 480(4), pp. 415–426.

McKusick, V. (1965) 'The Genetic Mucopolysaccharidoses', *Circulation*, 31(1).

Merry, C. L. R. *et al.* (1999) 'Highly sensitive sequencing of the sulfated domains of heparan sulfate', *Journal of Biological Chemistry*, 274(26), pp. 18455–18462.

Merry, C. L. R. *et al.* (2001) 'The Molecular Phenotype of Heparan Sulfate in the Hs2st-/- Mutant Mouse', *Journal of Biological Chemistry*, 276(38), pp. 35429–35434.

Merry, C. L. R. *et al.* (2022) 'Proteoglycans and Sulfated Glycosaminoglycans', *Essentials of Glycobiology*. Cold Spring Harbor Laboratory Press.

Merry, C. L. R. and Wilson, V. A. (2002) 'Role of heparan sulfate-2-O-sulfotransferase in the mouse', *Biochimica et Biophysica Acta - General Subjects*, 1573(3), pp. 319–327..

Mii, Y. and Takada, S. (2020) 'Heparan Sulfate Proteoglycan Clustering in Wnt Signaling and Dispersal', *Frontiers in Cell and Developmental Biology*, 8,(631).

Miyamoto, K. *et al.* (2003) 'Methylation-associated silencing of heparan sulfate D-glucosaminyl 3-O-sulfotransferase-2 (3-OST-2) in human breast, colon, lung and pancreatic cancers', *Oncogene*, 22(2), pp. 274–280.

Mizumoto, S. and Yamada, S. (2021) 'Congenital Disorders of Deficiency in Glycosaminoglycan Biosynthesis', *Frontiers in Genetics*, 12.

Mollaki, V. (2021) 'Ethical challenges in organoid use', *BioTech*, 10(3), pp. 12.

Moradi, S. *et al.* (2019) 'Research and therapy with induced pluripotent stem cells (iPSCs): Social, legal, and ethical considerations', *Stem Cell Research and Therapy*, 10(1), pp. 1–13.

Morimoto-Tomita, M. *et al.* (2002) 'Cloning and characterization of two extracellular heparin-degrading endosulfatases in mice and humans', *Journal of Biological Chemistry*, 277(51), pp. 49175–49185.

Mosqueira, D. *et al.* (2018) 'CRISPR/Cas9 editing in human pluripotent stemcell-cardiomyocytes highlights arrhythmias, hypocontractility, and energy depletion as potential therapeutic targets for hypertrophic cardiomyopathy', *European Heart Journal*, 39(43), pp. 3879–3892.

Mossahebi-Mohammadi, M. *et al.* (2020) 'FGF Signaling Pathway: A Key Regulator of Stem Cell Pluripotency', *Frontiers in Cell and Developmental Biology*, 8(79).

Mou, H. *et al.* (2017) 'CRISPR/Cas9-mediated genome editing induces exon skipping by alternative splicing or exon deletion', *Genome Biology*, 18(1), pp. 1–8.

Muenzer, J. (2011) 'Overview of the mucopolysaccharidoses', *Rheumatology*, 50(5), pp. 4–12.

Mulligan, K. A. and Cheyette, B. N. R. (2012) 'Wnt signaling in vertebrate neural development and function', *Journal of Neuroimmune Pharmacology*, 7(4), pp. 774–787.

Munji, R. N. *et al.* (2011) 'Wnt signaling regulates neuronal differentiation of cortical intermediate progenitors', *Journal of Neuroscience*, 31(5), pp. 1676–1687.

Nam, E. J. and Park, P. W. (2012) 'Shedding of cell membrane-bound proteoglycans', *Methods in Molecular Biology*, 836, pp. 291–305.

National Institute For Health and Care Excellence (2022) *Final evaluation determination - Atidarsagene autotemcel for treating metachromatic leukodystrophy*.

Nguyen, T. K. N. *et al.* (2012) 'A synthetic heparan sulfate oligosaccharide library reveals the novel enzymatic action of d-glucosaminyl 3-O-sulfotransferase-3a', *Molecular BioSystems*, 8(2), pp. 609–614.

Okajima, T. *et al.* (1999) 'Molecular basis for the progeroid variant of Ehlers-Danlos syndrome. Identification and characterization of two mutations in galactosyltransferase I gene', *Journal of Biological Chemistry*, 274(41), pp. 28841–28844.

Orchard Therapeutics (2022) *Orchard Therapeutics Announces Promising Early Neurocognitive Outcomes from Ongoing Proof-of-concept Study of OTL-201 in MPS-IIIA*. Available at: www.orchard-tx.com (Accessed: 28 February 2023).

Pallerla, S. R. *et al.* (2008) 'Altered heparan sulfate structure in mice with deleted NDST3 gene function', *Journal of Biological Chemistry*, 283(24), pp. 16885–16894.

Pan, C. *et al.* (2005) 'Functional abnormalities of heparan sulfate in mucopolysaccharidosis-I are associated with defective biologic activity of FGF-2 on human multipotent progenitor cells', *Blood*, 106(6), pp. 1956–1964.

Passarelli, M. K. *et al.* (2017) 'The 3D OrbiSIMS - Label-free metabolic imaging with subcellular lateral resolution and high mass-resolving power', *Nature Methods*, 14(12), pp. 1175–1183.

Paul, P. *et al.* (2012) 'Recent advances in sulfotransferase enzyme activity assays', *Analytical and Bioanalytical Chemistry*, 403(6), pp. 1491–1500.

Pera, E. M. *et al.* (2003) 'Integration of IGF, FGF, and anti-BMP signals via Smad1 phosphorylation in neural induction', *Genes and Development*, 17(24), pp. 3023–3028.

Pickford, C. E. *et al.* (2011) 'Specific glycosaminoglycans modulate neural specification of mouse embryonic stem cells', *Stem Cells*, 29(4), pp. 629–640.

Pinhal, M. A. S. *et al.* (2001) Enzyme interactions in heparan sulfate biosynthesis: Uronosyl 5-epimerase and 2-O-sulfotransferase interact in vivo, *Proceedings of the National Academy of Sciences of the United States of America*, 98(23), pp.12984-12989.

Poletto, E., Baldo, G. and Gomez-Ospina, N. (2020) 'Genome editing for mucopolysaccharidoses', *International Journal of Molecular Sciences*, 21(2), pp. 500.

Pye, D. A. *et al.* (1998) 'Heparan sulfate oligosaccharides require 6-O-sulfation for promotion of basic fibroblast growth factor mitogenic activity', *Journal of Biological Chemistry*, 273(36), pp. 22936–22942.

Qiu, H. *et al.* (2018) 'A mutant-cell library for systematic analysis of heparan sulfate structure–function relationships', *Nature Methods*, 15(11), pp. 889–899.

Quentin, E. *et al.* (1990) 'A genetic defect in the biosynthesis of dermatan sulfate proteoglycan: Galactosyltransferase I deficiency in fibroblasts from a patient with a progeroid syndrome', Proceedings of the National Academy of Sciences, 87(4), pp. 1342–1346.

Ratajczak, M. Z. *et al.* (2012) 'Pluripotent and multipotent stem cells in adult tissues', *Advances in Medical Sciences*. Elsevier, 57(1), pp. 1–17.

Reinders, J. and Sickmann, A. (2005) 'State-of-the-art in phosphoproteomics', *Proteomics*, 5(16), pp. 4052–4061.

Reuter, M. S. *et al.* (2014) 'NDST1 missense mutations in autosomal recessive intellectual disability', *American Journal of Medical Genetics Part A*, 164(11), pp. 2753–2763.

Rice, K. G. *et al.* (1985) 'High-performance liquid chromatographic separation of heparin-derived oligosaccharides', *Analytical Biochemistry*, 150(2), pp. 325–331.

Rider, C. (1998) 'Analysis of glycosaminoglycans and proteoglycans.', *Glycoanalysis Protocols*, 76, pp. 131–143.

Ringvall, M. *et al.* (2000) 'Defective heparan sulfate biosynthesis and neonatal lethality in mice lacking N-deacetylase/N-sulfotransferase-1', *Journal of Biological Chemistry*, 275(34), pp. 25926–25930.

Rossi, A. and Brunetti-Pierri, N. (2023) 'Gene therapies for mucopolysaccharidoses', *Journal of Inherited Metabolic Disease*.

Roughley, P. J. and Mort, J. S. (2014) 'The role of aggrecan in normal and osteoarthritic cartilage', *Journal of Experimental Orthopaedics*, 1(1), pp. 1–11.

Ruiz i Altaba, A. (1994) 'Pattern formation in the vertebrate neural plate', *Trends in Neurosciences*, 17(6), pp. 233–243.

Sahabian, A. *et al.* (2021) 'Production and cryopreservation of definitive endoderm from human pluripotent stem cells under defined and scalable culture conditions', *Nature Protocols*, 16(3), pp. 1581–1599.

Sander, J. D. and Joung, J. K. (2014) 'CRISPR-Cas systems for editing, regulating and targeting genomes', *Nature Biotechnology*, 32(4), pp. 347–350.

Sasai, Y. *et al.* (1995) 'Regulation of neural induction by the Chd and Bmp-4 antagonistic patterning signals in Xenopus', *Nature*, 376(6538), pp. 333–336.

Sasai, Y. and De Robertis, E. M. (1997) 'Ectodermal patterning in vertebrate embryos', *Developmental Biology*, 182(1), pp. 5–20.

Schneeberger, P. E. *et al.* (2020) 'Bi-allelic Pathogenic Variants in HS2ST1 Cause a Syndrome Characterized by Developmental Delay and Corpus Callosum, Skeletal, and Renal Abnormalities', *American Journal of Human Genetics*, 107(6), pp. 1044–1061.

Schuchardt, A. *et al.* (1994) 'Defects in the kidney and enteric nervous system of mice lacking the tyrosine kinase receptor Ret', *Nature*, 367(6461), pp. 380–383.

Seker Yilmaz, B. *et al.* (2021) 'Novel therapies for mucopolysaccharidosis type III', *Journal of Inherited Metabolic Disease*, 44(1), pp. 129–147.

Sepulveda-Diaz, J. E. *et al.* (2015) 'HS3ST2 expression is critical for the abnormal phosphorylation of tau in Alzheimer's disease-related tau pathology', *Brain*, 128, pp. 1339–1354.

Shard, A. G. *et al.* (2015) 'The matrix effect in organic secondary ion mass spectrometry', *International Journal of Mass Spectrometry*, 377(1), pp. 599–609.

Sharkia, R. *et al.* (2014) 'Sanfilippo type A: New clinical manifestations and neuro-imaging findings in patients from the same family in Israel: A case report', *Journal of Medical Case Reports*, 8(1), pp. 1–6.

Shworak, N. W. *et al.* (1996) 'Cell-free synthesis of anticoagulant heparan sulfate reveals a limiting converting activity that modifies an excess precursor pool', *Journal of Biological Chemistry*, 271(43), pp. 27063–27071.

Shworak, N. W. *et al.* (1999) 'Multiple Isoforms of Heparan Sulfate d-Glucosaminyl 3-O-Sulfotransferase', *Journal of Biological Chemistry*,

274(8), pp. 5170–5184.

Shworak, N. W. *et al.* (2002) 'Mice deficient in heparan sulfate 3-O-sulfotransferase-1: Normal hemostasis with unexpected perinatal phenotypes', *Glycoconjugate Journal*, 19(4–5), pp. 355–361.

Silver, F. H., Kelkar, N. and Deshmukh, T. (2021) 'Molecular basis for mechanical properties of ECMs: Proposed role of fibrillar collagen and proteoglycans in tissue biomechanics', *Biomolecules*, 11(7), pp. 1018.

Singh, A. M. *et al.* (2012) 'Signaling network crosstalk in human pluripotent cells: A Smad2/3-regulated switch that controls the balance between self-renewal and differentiation', *Cell Stem Cell*, 10(3), pp. 312–326.

Smeds, E. *et al.* (2003) 'Substrate specificities of mouse heparan sulphate glucosaminyl 6-O-sulphotransferases', *Biochemical Journal*, 372(2), pp. 371–380.

Smits, N. C. *et al.* (2006) 'Phage Display-Derived Human Antibodies Against Specific Glycosaminoglycan Epitopes', *Methods in Enzymology*, 416, pp. 61–87.

Smits, N. C. *et al.* (2010) 'The heparan sulfate motif (GlcNS6S-IdoA2S)3, common in heparin, has a strict topography and is involved in cell behavior and disease', *Journal of Biological Chemistry*, 285(52), pp. 41143–41151.

Souza-Fernandes, A. B., Pelosi, P. and Rocco, P. R. M. (2006) 'Bench-to-bedside review: The role of glycosaminoglycans in respiratory disease', *Critical Care*, 10(6), pp. 237.

Spemann, H. and Mangold, H. (2001) 'Induction of embryonic primordia by implantation of organizers from a different species', *International Journal of Developmental Biology*, 45(1), pp. 13–38.

Stauffer, S., Biswas, K. and Sharan, S. K. (2020) 'Bypass of premature stop codons and generation of functional BRCA2 by exon skipping', *Journal of Human Genetics*, 65(9), pp. 805–809.

Stenson, P. D. *et al.* (2003) 'Human Gene Mutation Database (HGMD®): 2003 Update', *Human Mutation*, 21(6), pp. 577–581.

Stepp, M. A. *et al.* (2015) 'Syndecan-1 and Its Expanding List of Contacts', *Advances in Wound Care*, 4(4), pp. 235–249.

Stern, C. D. (2005) 'Neural induction: Old problem, new findings, yet more questions', *Development*, 132(9), pp. 2007–2021.

Sterner, E. *et al.* (2014) 'Assays for determining heparan sulfate and heparin O-sulfotransferase activity and specificity', *Analytical and Bioanalytical Chemistry*, 406(2), pp. 525–536.

Stickens, D. *et al.* (2005) 'Mice deficient in Ext2 lack heparan sulfate and develop exostoses', *Development*, 133(22), pp. 5055–5068.

Sugahara, K. and Kitagawa, H. (2002) 'Heparin and heparan sulfate biosynthesis', *IUBMB Life*, 54(4), pp. 163–175.

Sugaya, N. *et al.* (2008) '6-O-sulfation of heparan sulfate differentially regulates various fibroblast growth factor-dependent signalings in

culture', *Journal of Biological Chemistry*, 283(16), pp. 10366–10376.

Syx, D. *et al.* (2022) 'Alterations in glycosaminoglycan biosynthesis associated with the Ehlers-Danlos syndromes', *American journal of physiology. Cell physiology*, 323(6), pp. C1843–C1859.

Takahashi, K. and Yamanaka, S. (2006) 'Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors', *Cell*, 126(4), pp. 663–676.

Takata, A. *et al.* (2011) 'Direct differentiation of hepatic cells from human induced pluripotent stem cells using a limited number of cytokines', *Hepatology International*, 5(4), pp. 890–898.

Tanabe, H., Yokota, K. and Kohgo, Y. (1999) 'Localization of syndecan-1 in human gastric mucosa associated with ulceration', *Journal of Pathology*, 187(3), pp. 338–344.

Thacker, B. E. *et al.* (2014) 'Heparan sulfate 3-O-sulfation: A rare modification in search of a function', *Matrix Biology*, 35, pp. 60–72.

The, I., Bellaiche, Y. and Perrimon, N. (1999) 'Hedgehog movement is regulated through tout velu-dependent synthesis of a heparan sulfate proteoglycan', *Molecular Cell*, 4(4), pp. 633–639.

Thermo Fisher Scientific Inc. (2015) *Differentiating Neural Stem Cells into Neurons and Glial Cells*, *Neurobiology Protocols*. Available at: https://www.thermofisher.com/uk/en/home/references/protocols/neurobiology/neurobiology-protocols/differentiating-neural-stem-cells-into-neurons-and-glial-cells.html (Accessed: 13 March 2023).

Thompson, S. M. *et al.* (2009) 'Heparan sulfate phage display antibodies identify distinct epitopes with complex binding characteristics: Insights into protein binding specificities', *Journal of Biological Chemistry*, 284(51), pp. 35621–35631.

Thomson, J. A. *et al.* (1998) 'Embryonic stem cell lines derived from human blastocysts', *Science*, 6 November, pp. 1145–1147.

Tkachenko, E., Rhodes, J. M. and Simons, M. (2005) 'Syndecans: New kids on the signaling block', *Circulation Research*, 96(5), pp. 488–500.

Toma, L., Berninsone, P. and Hirschberg, C. B. (1998) The putative heparin-specific N-acetylglucosaminyl N-deacetylase/N- sulfotransferase also occurs in non-heparin-producing cells, *Journal of Biological Chemistry*, 273(35), pp.22458-22465.

Tone, Y. *et al.* (2008) '2-O-phosphorylation of xylose and 6-O-sulfation of galactose in the protein linkage region of glycosaminoglycans influence the glucuronyltransferase-I activity involved in the linkage region synthesis', *Journal of Biological Chemistry*, 283(24), pp. 16801–16807.

Tornberg, J. *et al.* (2011) 'Heparan sulfate 6-O-sulfotransferase 1, a gene involved in extracellular sugar modifications, is mutated in patients with idiopathic hypogonadotrophic hypogonadism', Proceedings of the National Academy of Sciences of the United States of America, 108(28), pp. 11524–11529.

Toydemir, R. M. *et al.* (2006) 'A novel mutation in FGFR3 causes camptodactyly, tall stature, and hearing loss (CATSHL) syndrome',

American Journal of Human Genetics, 79(5), pp. 935–941.

Turnbull, J. E. *et al.* (1992) 'Identification of the basic fibroblast growth factor binding sequence in fibroblast heparan sulfate', *Journal of Biological Chemistry*, 267(15), pp. 10337–10341.

Uhl, E. W. and Warner, N. J. (2015) 'Mouse Models as Predictors of Human Responses: Evolutionary Medicine', *Current Pathobiology Reports*, 3(3), pp. 219–223.

Vallejo, S. *et al.* (2018) 'Generation of two induced pluripotent stem cells lines from Mucopolysaccharydosis IIIA patient: IMEDEAi004-A and IMEDEAi004-B', *Stem Cell Research*, 32, pp. 110–114.

Vallier, L. *et al.* (2009) 'Activin/Nodal signalling maintains pluripotency by controlling Nanog expression', *Development*, 136(8), pp. 1339–1349.

Vallier, L., Alexander, M. and Pedersen, R. A. (2005) 'Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells', *Journal of Cell Science*, 118(19), pp. 4495–4509.

Vallier, L., Reynolds, D. and Pedersen, R. A. (2004) 'Nodal inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway', *Developmental Biology*, 275(2), pp. 403–421.

Valstar, M. J. *et al.* (2008) 'Sanfilippo syndrome: A mini-review', *Journal of Inherited Metabolic Disease*, 31(2), pp. 240–252.

Valstar, M. J. *et al.* (2010) 'Mucopolysaccharidosis type IIIA: Clinical spectrum and genotype-phenotype correlations', *Annals of Neurology*, 68(6), pp. 876–887.

Velleman, S. G. and Song, Y. (2017) 'Development and growth of the avian pectoralis major (Breast) muscle: Function of syndecan-4 and glypican-1 in adult myoblast proliferation and differentiation', *Frontiers in Physiology*. Frontiers Media S.A., 8, pp. 577.

Veraldi, N. *et al.* (2018) 'Structural features of heparan sulfate from multiple osteochondromas and chondrosarcomas', *Molecules*, 23(12), pp. 3277.

Viana, G. M. *et al.* (2020) 'Brain pathology in mucopolysaccharidoses (MPS) patients with neurological forms', *Journal of Clinical Medicine*, 9(2), pp. 396.

Vickerman, J. C. (2013) '1 - Prologue: ToF-SIMS - An evolving mass spectrometry of materials', in Vickerman, J. C. and Briggs, D. (eds) *TOF-SIMS: Materials Analysis by Mass Spectrometry*. 2nd edn. IM Publications LLP and Surface Spectra Limited, pp. 1–37.

Volpi, N. and Linhardt, R. J. (2010) 'High-performance liquid chromatography-mass spectrometry for mapping and sequencing glycosaminoglycan-derived oligosaccharides', *Nature Protocols*, 5(6), pp. 993–1004.

Vreys, V. and David, G. (2007) 'Mammalian heparanase: what is the message?', *Journal of cellular and molecular medicine*, 11(3), pp. 427–452.

Walkley, S. U. (2004) 'Secondary accumulation of gangliosides in

lysosomal storage disorders', *Seminars in Cell and Developmental Biology*, 15(4), pp. 433–444.

Walma, D. A. C. and Yamada, K. M. (2020) 'The extracellular matrix in development', *Development*, 147(10).

Wang, H. *et al.* (1997) 'Glycosaminoglycans can influence fibroblast growth factor-2 mitogenicity without significant growth factor binding', *Biochemical and Biophysical Research Communications*, 235(2), pp. 369–373.

Wang, L. *et al.* (2019) 'Delivery of CRISPR/Cas9 by Novel Strategies for Gene Therapy', *ChemBioChem*, 20(5), pp. 634–643.

Wang, Xiaofeng *et al.* (2015) 'FGFR3/fibroblast growth factor receptor 3 inhibits autophagy through decreasing the ATG12–ATG5 conjugate, leading to the delay of cartilage development in achondroplasia', *Autophagy*, 11(11), pp. 1998–2013.

Wells, J. M. and Melton, D. A. (1999) *Vertebrate endoderm development*, *Annual Review of Cell and Developmental Biology*, 15(1), pp.393-410.

Wen, J. *et al.* (2008) 'Pax6 directly modulate Sox2 expression in the neural progenitor cells', *NeuroReport*, 19(4), pp. 413–417.

Wen, X. *et al.* (2016) 'Chondrocyte FGFR3 regulates bone mass by inhibiting osteogenesis', *Journal of Biological Chemistry*, 291(48), pp. 24912–24921.

Wijnhoven, T. J. M. *et al.* (2006) 'Aberrant Heparan Sulfate Profile in the Human Diabetic Kidney Offers New Clues for Therapeutic Glycomimetics', *American Journal of Kidney Diseases*, 48(2), pp. 250–261.

Wilkinson, F. L. *et al.* (2012) 'Neuropathology in mouse models of mucopolysaccharidosis type I, IIIA and IIIB', *PLoS ONE*, 7(4).

Wu, S. M. and Hochedlinger, K. (2011) 'Harnessing the potential of induced pluripotent stem cells for regenerative medicine', *Nature Cell Biology*, 13(5), pp. 497–505.

Wu, Z. L. *et al.* (2004) 'Determining Heparan Sulfate Structure in the Vicinity of Specific Sulfotransferase Recognition Sites by Mass Spectrometry', *Journal of Biological Chemistry*, 279(3), pp. 1861–1866.

Wuechner, C. *et al.* (1996) 'Developmental expression of splicing variants of fibroblast growth factor receptor 3 (FGFR3) in mouse', *International Journal of Developmental Biology*, 40(6), pp. 1185–1188.

Wyatt, D. and Ramsden, D. (2015) *CRISPR 101: Non-Homologous End Joining*, *Addgene Blog*. Available at: https://blog.addgene.org/crispr-101-non-homologous-end-joining (Accessed: 25 January 2022).

Xie, Y. *et al.* (2020) 'FGF/FGFR signaling in health and disease', *Signal Transduction and Targeted Therapy*, 5(1), pp. 1–38.

Xu, D. *et al.* (2005) 'Characterization of heparan sulphate 3-0-sulphotransferase isoform 6 and its role in assisting the entry of herpes simplex virus type 1', *Biochemical Journal*, 385(2), pp. 451–459.

Xu, X. *et al.* (2019) 'Delivery of CRISPR/Cas9 for therapeutic genome editing', *Journal of Gene Medicine*, 21(7).

Yamanaka, S. (2007) 'Strategies and New Developments in the Generation of Patient-Specific Pluripotent Stem Cells', *Cell Stem Cell*, 1(1), pp. 39–49..

Yang, B. *et al.* (2011) 'Hyphenated techniques for the analysis of heparin and heparan sulfate', *Analytical and Bioanalytical Chemistry*, 399(2), pp. 541–557.

Zhang, L., Beeler, D. L., *et al.* (2001) '6-O-Sulfotransferase-1 Represents a Critical Enzyme in the Anticoagulant Heparan Sulfate Biosynthetic Pathway', *The Journal of Biological Chemistry*, 276(45), pp. 42311–42321.

Zhang, L., Lawrence, R., *et al.* (2001) 'The Effect of Precursor Structures on the Action of Glucosaminyl 3-O-Sulfotransferase-1 and the Biosynthesis of Anticoagulant Heparan Sulfate', *Journal of Biological Chemistry*, 276(31), pp. 28806–28813.

Zhang, S. *et al.* (2020) 'Strategies in the delivery of Cas9 ribonucleoprotein for CRISPR/Cas9 genome editing', *Theranostics*, 11(2), pp. 614–648.

Zhang, Y. *et al.* (2013) 'Rapid single-step induction of functional neurons from human pluripotent stem cells', *Neuron*, 78(5), pp. 785–798.

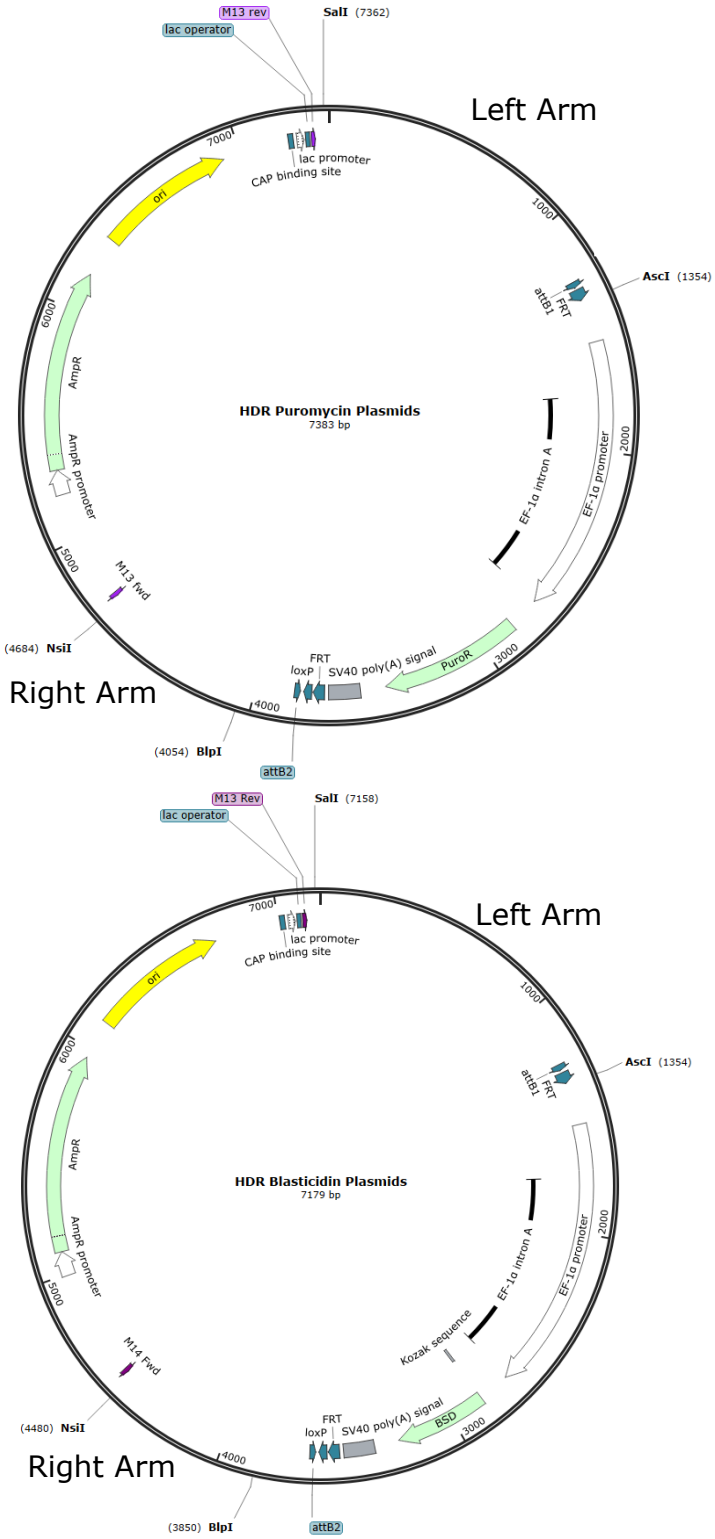
Zhao, J. *et al.* (2020) '3-O-Sulfation of Heparan Sulfate Enhances Tau Interaction and Cellular Uptake', *Angewandte Chemie - International Edition*, 59(5), pp. 1818–1827.

Zheng, Y. L. (2016) 'Some Ethical Concerns About Human Induced Pluripotent Stem Cells', *Science and Engineering Ethics*, 22(5), pp. 1277–1284.

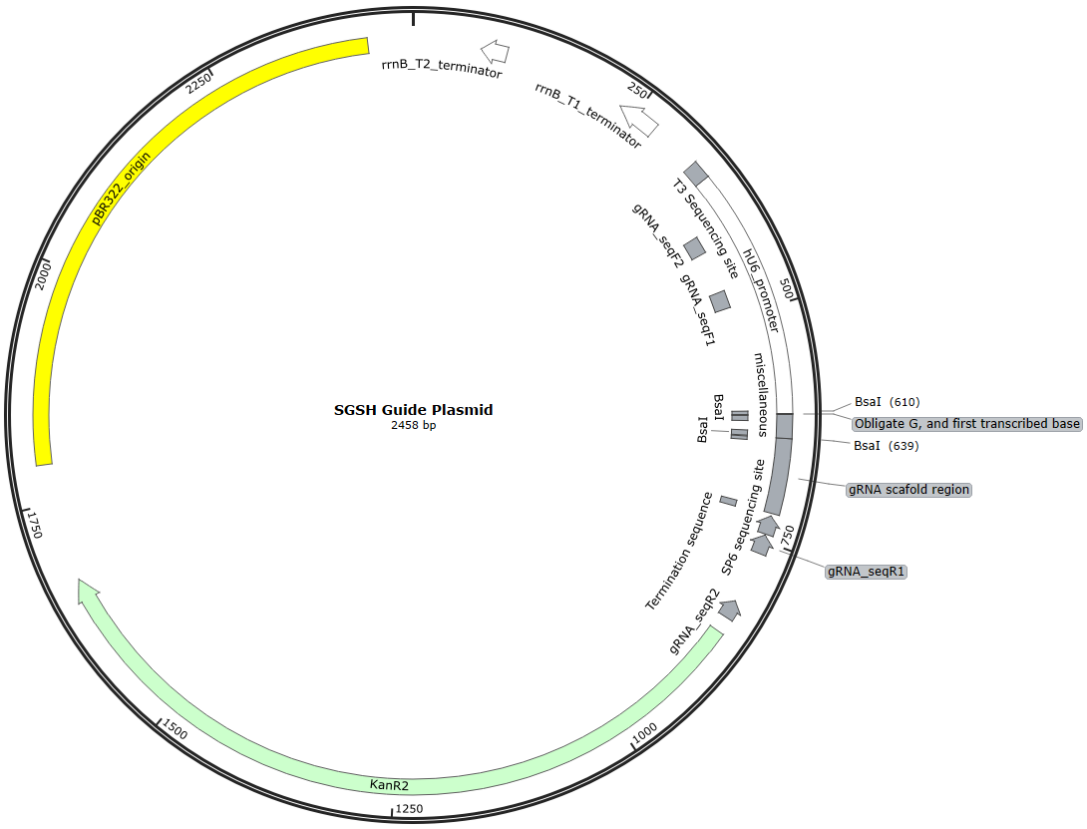
Zhou, T. *et al.* (2011) 'Generation of induced pluripotent stem cells from urine', *Journal of the American Society of Nephrology*, 22(7), pp. 1221–1228.

Zorn, A. M. and Wells, J. M. (2009) 'Vertebrate endoderm development and organ formation', *Annual Review of Cell and Developmental Biology*, 25, pp. 221–251.

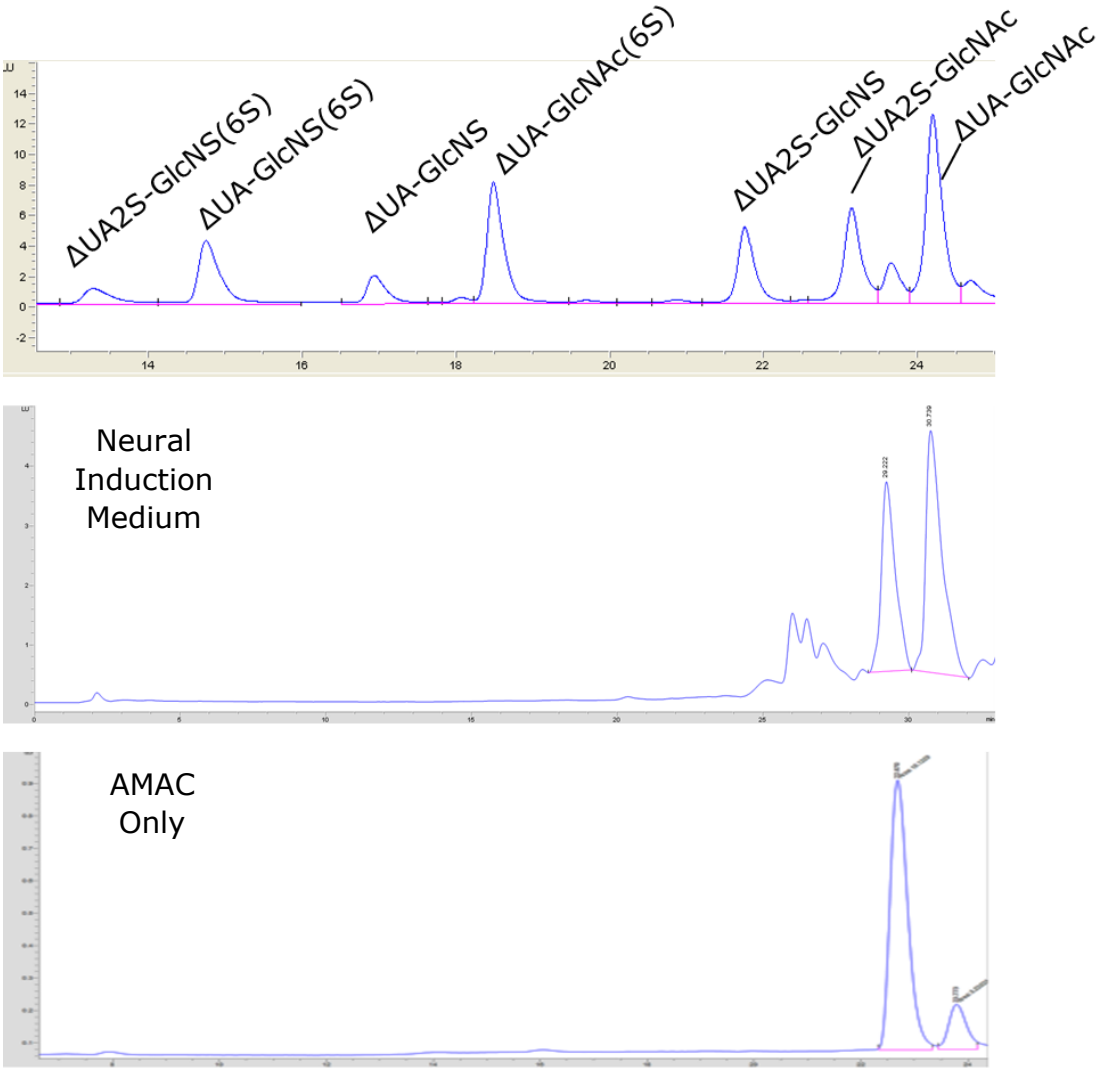
Appendix



Appendix 1 – CRISPR/Cas9 HDR Plasmids. Schematics of *SGSH* HDR targeting plasmids with labelled restriction sites (SalI, AscI, BlpI and NsiI). These plasmids carry antibiotic resistance cassettes for A) puromycin and B) blasticidin. These plasmids would have the arms of homology inserted between enzyme restriction sides at noted on each plasmid.



Appendix 2 – CRISPR/Cas9 Guide Plasmid. Schematic of *SGSH* targeting plasmids that contains the guide sequence that undergoes transcription into gRNA and form a complex with the Cas9. The gRNA directs the Cas9 to the correct gene locus.



Appendix 3 - Neural Induction Medium is Heparan Sulphate Free. AMAC labelling and RP-HPLC were used to confirm that neural induction supplement was HS-free. A) HS disaccharide standards B) neural induction medium c) AMAC only